



Statistical Policy
Working Paper 20

Seminar on Quality of Federal Data

Part 2 of 3

Federal Committee on Statistical Methodology

Statistical Policy Office
Office of Information and Regulatory Affairs
Office of Management and Budget

March 1991

DID 416
C-10 170

**MEMBERS OF THE FEDERAL COMMITTEE ON
STATISTICAL METHODOLOGY**

(February 1991)

**Maria E. Gonzalez, Chair
Office of Management and Budget**

**Yvonne M. Bishop
Energy Information
Administration**

**Warren L. Buckler
Social Security Administration**

**Charles E. Caudill
National Agricultural
Statistics Service**

**Cynthia Z.F. Clark
National Agricultural
Statistics Service**

**Zahava D. Doering
Smithsonian Institution**

**Robert M. Groves
Bureau of the Census**

**Roger A. Herriot
National Center for
Education Statistics**

**C. Terry Ireland
National Computer Security
Center**

**Charles D. Jones
Bureau of the Census**

**Daniel Kasprzyk
Bureau of the Census**

**Daniel Melnick
National Science Foundation**

**Robert P. Parker
Bureau of Economic Analysis**

**David A. Pierce
Federal Reserve Board**

**Thomas J. Plewes
Bureau of Labor Statistics**

**Wesley L. Schaible
Bureau of Labor Statistics**

**Fritz J. Scheuren
Internal Revenue Service**

**Monroe G. Sirken
National Center for
Health Statistics**

**Robert D. Tortora
Bureau of the Census**

PREFACE

In 1975, the Office of Management and Budget (OMB) organized the Federal Committee on Statistical Methodology. Comprised of individuals selected by OMB for their expertise and interest in statistical methods, the committee has during the past 15 years determined areas that merit investigation and discussion, and overseen the work of subcommittees organized to study particular issues. Since 1978, 19 Statistical Policy Working Papers have been published under the auspices of the Committee.

On May 23-24, 1990, the Council of Professional Associations on Federal Statistics (COPAFS) hosted a "Seminar on the Quality of Federal Data." Developed to capitalize on work undertaken during the past dozen years by the Federal Committee on Statistical Methodology and its subcommittees, the seminar focused on a variety of topics that have been explored thus far in the Statistical Policy Working Paper series. The subjects covered at the seminar included:

- Survey Quality Profiles
- Paradigm Shifts Using Administrative Records
- Survey Coverage Evaluation
- Telephone Data Collection
- Data Editing
- Computer Assisted Statistical Surveys
- Quality in Business Surveys
- Cognitive Laboratories
- Employer Reporting Unit Match Study
- Approaches to Developing Questionnaires
- Statistical Disclosure-Avoidance
- Federal Longitudinal Surveys

Each of these topics was presented in a two-hour session that featured formal papers and discussion, followed by informal dialogue among all speakers and attendees.

Statistical Policy Working Paper 20, published in three parts, presents the proceedings of the "Seminar on the Quality of Federal Data." In addition to providing the papers and formal discussions from each of the twelve sessions, this working paper includes Robert M. Groves' keynote address, "Towards Quality in a Working Paper Series on Quality," and comments by Stephen E. Fienberg, Margaret E. Martin, and Hermann Habermann at the closing session, "Towards an Agenda for the Future."

We are indebted to all of our colleagues who assisted in organizing the seminar, and to the many individuals who not only presented papers and discussions but also prepared these materials for publication. A special thanks is due to Terry Ireland and his staff for their work in assembling this working paper.

Table of Contents

Wednesday, May 23, 1990

Part 1

KEYNOTE ADDRESS

TOWARDS QUALITY IN A WORKING PAPER SERIES ON QUALITY.	3
Robert M. Groves, The University of Michigan and U. S. Bureau of the Census	

Session 1 - SURVEY QUALITY PROFILES

THE SIPP QUALITY PROFILE.	19
Thomas B. Jabine, Statistical Consultant	
INITIAL REPORT ON THE QUALITY OF AGRICULTURAL SURVEY PROGRAM. .	29
George A. Hanuschak, National Agricultural Statistics Service	
DISCUSSION.	40
Barbara A. Bailar, American Statistical Association	
DISCUSSION.	46
Nancy A. Mathiowetz, U. S. Bureau of the Census	

Session 2 - PARADIGM SHIFTS USING ADMINISTRATIVE RECORDS

PARADIGM SHIFTS: ADMINISTRATIVE RECORDS AND CENSUS-TAKING. . .	53
Fritz Scheuren, Internal Revenue Service	
AN ADMINISTRATIVE RECORD PARADIGM: A CANADIAN EXPERIENCE . . .	66
John Leyes, Statistics Canada	

DISCUSSION.	77
Gerald Gates, U.S. Bureau of the Census	
DISCUSSION.	83
Edward J. Spar, Market Statistics	

Session 3 - SURVEY COVERAGE EVALUATION

CONTROL MEASUREMENT, AND IMPROVEMENT OF SURVEY COVERAGE	87
Gary M. Shapiro, U. S. Bureau of the Census; Raymond R. Bosecker, National Agricultural Statistics Service	
QUALITY OF SURVEY FRAMES.	100
Judith T. Lessler, Research Triangle Institute	
DISCUSSION.	108
Fritz Scheuren, Internal Revenue Service	
DISCUSSION.	114
Joseph Waksberg, Westat, Inc.	

Session 4 - TELEPHONE DATA COLLECTION

QUALITY IMPROVEMENT IN TELEPHONE SURVEYS.	123
Leyla Mohadjer, David Morganstein, Westat, Inc.	
COMPUTER ASSISTED SURVEY TECHNOLOGIES IN GOVERNMENT:	
AN OVERVIEW.	137
Marc Tosiano, National Agricultural Statistics Service	
DISCUSSION.	155
William L. Nicholls II, U. S. Bureau of the Census	
DISCUSSION.	161
James T. Massey, National Center for Health Statistics	

Part 2

Session 5 - DATA EDITING

OVERVIEW OF DATA EDITING IN FEDERAL STATISTICAL AGENCIES. . .	.167
David A. Pierce, Federal Reserve Board	
EDITING SOFTWARE (An excerpt from Chapter IV of Working Paper 18).173
Mark Pierzchala, National Agricultural Statistics Service	
RESEARCH ON EDITING180
Yahia Ahmed, Internal Revenue Service	
DISCUSSION.184
Charles E. Caudill, National Agricultural Statistics Service	
DISCUSSION.186
Richard Bolstein, George Mason University	

Session 6 - COMPUTER ASSISTED STATISTICAL SURVEYS

OVERVIEW OF COMPUTER ASSISTED SURVEY INFORMATION COLLECTION .	.191
Richard L. Clayton, U. S. Bureau of Labor Statistics	
A COMPARISON BETWEEN CATI AND CAPI.197
Martin Baum, National Center for Health Statistics	
COMPUTER ASSISTED SELF INTERVIEWING202
Ralph Gillmann, Energy Information Administration	
COMPUTER ASSISTED SELF INTERVIEWING: RIGS AND PEDRO, TWO EXAMPLES205
Ann M. Ducca, Energy Information Administration	
DATA COLLECTION209
Cathy Mazur, National Agricultural Statistics Service	

DISCUSSION.212
Robert N. Tinari, U. S. Bureau of the Census	
DISCUSSION.216
David Morganstein, Westat, Inc.	

Thursday, May 24, 1990

Session 7 - QUALITY IN BUSINESS SURVEYS

IMPROVING ESTABLISHMENT SURVEYS AT THE BUREAU OF LABOR STATISTICS221
Brian MacDonald, Alan R. Tupek, U. S. Bureau of Labor Statistics	
A REVIEW OF NONSAMPLING ERRORS IN FEDERAL ESTABLISHMENT SURVEYS WITH SOME AGRIBUSINESS EXAMPLES232
Ron Fecso, National Agricultural Statistics Service	
DISCUSSION.243
David A. Binder, Statistics Canada	
DISCUSSION.247
Charles D. Cowan, Opinion Research Corporation	

Session 8 - COGNITIVE LABORATORIES

THE BUREAU OF LABOR STATISTICS' COLLECTION PROCEDURES RESEARCH LABORATORY: ACCOMPLISHMENTS AND FUTURE DIRECTIONS . .	.253
Cathryn S. Dippe, Douglas Herrmann, U. S. Bureau of Labor Statistics	
THE ROLE OF A COGNITIVE LABORATORY IN A STATISTICAL AGENCY. .	.268
Monroe G. Sirken, National Center for Health Statistics	
DISCUSSION.278
Elizabeth Martin, U. S. Bureau of the Census	
DISCUSSION.281
Murray Aborn, National Science Foundation (retired)	

Part 3

Session 9 - EMPLOYER REPORTING UNIT MATCH STUDY

INTERAGENCY AGREEMENTS FOR MICRODATA ACCESS:	
THE ERUMS EXPERIENCE291
Thomas B. Petska, Internal Revenue Service; Lois Alexander, Social Security Administration	
SAMPLE SELECTION AND MATCHING PROCEDURES USED IN ERUMS.301
John Pinkos, Kenneth LeVasseur, Marlene Einstein, U. S. Bureau of Labor Statistics; Joel Packman, Social Security Administration	
RESULTS, FINDINGS, AND RECOMMENDATIONS OF THE ERUMS PROJECT309
Vern Renshaw, Bureau of Economic Analysis; Tom Jabine, Statistical Consultant	
DISCUSSION.318
W. Joel Richardson, Charles A. Waite, U. S. Bureau of the Census	
DISCUSSION.324
Thomas J. Plewes, U. S. Bureau of Labor Statistics	

Session 10 - APPROACHES TO DEVELOPING QUESTIONNAIRES

TOOLS FOR USE IN DEVELOPING QUESTIONS AND TESTING QUESTIONNAIRES331
Theresa J. DeMaio, U. S. Bureau of the Census	
TECHNIQUES FOR EVALUATING THE QUESTIONNAIRE DRAFT340
Deborah H. Bercini, National Center for Health Statistics	
DESIGNING QUESTIONNAIRES FOR CATI IN A MIXED MODE ENVIRONMENT.349
Gemma Furno, U. S. Bureau of the Census	
DISCUSSION.360
Carol C. House, National Agricultural Statistics Service	

Session 11 - STATISTICAL DISCLOSURE - AVOIDANCE

DISCLOSURE AVOIDANCE PRACTICES AT THE CENSUS BUREAU367
Brian Greenberg, U. S. Bureau of the Census	
THE MICRODATA RELEASE PROGRAM OF THE NATIONAL CENTER FOR HEALTH STATISTICS377
Robert H. Mugge, National Center for Health Statistics (retired)	
DISCUSSION.385
George T. Duncan, Carnegie Mellon University	

Session 12 - FEDERAL LONGITUDINAL SURVEYS

FEDERAL LONGITUDINAL SURVEYS.393
Daniel Kasprzyk, U. S. Bureau of the Census; Curtis Jacobs, U. S. Bureau of Labor Statistics	
THE ADVANTAGES AND DISADVANTAGES OF LONGITUDINAL SURVEYS.407
Robert W. Pearson, Social Science Research Council	
LONGITUDINAL ANALYSIS OF FEDERAL SURVEY DATA.425
Patricia Ruggles, Joint Economic Committee	
DISCUSSION.438
Michael Brick, Westat, Inc.	
DISCUSSION.447
Marilyn E. Manser, U. S. Bureau of Labor Statistics	

TOWARDS AN AGENDA FOR THE FUTURE

Stephen E. Fienberg, Carnegie Mellon University455
Margaret E. Martin.462
Hermann Habermann, Office of Management and Budget.465

Part 2
Session 5
DATA EDITING

OVERVIEW OF DATA EDITING IN FEDERAL STATISTICAL AGENCIES

David A. Pierce¹
Federal Reserve Board

Abstract

This paper is the first of three in the session on Data Editing presenting highlights of the report "Data Editing in Federal Statistical Agencies", Statistical Policy Working Paper 18, OMB, prepared by the Subcommittee on Data Editing in Federal Statistical Agencies, FCSM. Included in this paper are a listing of the Subcommittee members, a discussion of its mission statement from the FCSM, definition and concepts of data editing, the major areas investigated and the methods used to do so, the development of case studies, and the Subcommittee's recommendations for data editing in Federal statistical agencies. The paper highlights the findings from a survey of current data editing practices which was conducted by the Subcommittee.

1. Introduction

The Subcommittee on Data Editing in Federal Statistical Agencies was established by the Federal Committee on Statistical Methodology (FCSM) in November 1988 to document, profile, and discuss the topic of data editing in Federal censuses and surveys. The Subcommittee consisted of the following individuals:

George Hanuschak, National Agricultural Statistics Service,
Chair
Yahia Ahmed, Internal Revenue Service
Laura Bauer, Federal Reserve Board
Charles Day, Internal Revenue Service
Maria Gonzalez, Office of Management and Budget
Brian Greenberg, Bureau of the Census
Anne Hafner, National Center for Education Statistics
Gerry Hendershot, National Center for Health Statistics
Rita Hohenbrink, National Agricultural Statistics Service
Renee Miller, Energy Information Administration
Tom Petkunas, Bureau of the Census
David Pierce, Federal Reserve Board

¹David A. Pierce is Senior Statistician, Micro Statistics Section, Division of Research and Statistics, Federal Reserve Board, Washington, DC 20551, and a member of the Federal Committee on Statistical Methodology and its Subcommittee on Data Editing in Federal Statistical Agencies. Any views expressed do not necessarily reflect those of the Federal Reserve System.

Mark Pierzchala, National Agricultural Statistics Service
Marybeth Tschetter, Bureau of Labor Statistics
Paula Weir, Energy Information Administration

A key aim of this effort was to further the awareness within agencies of each other's data editing practices, as well as of the state of the art of data editing, and thus to promote improvements in data quality throughout Federal statistical agencies. To further these goals, the Subcommittee was given a "charge", or mission statement, of

determining how data editing is currently being done in Federal agencies, recognizing areas that may need attention, and, if appropriate, recommending any potential improvements for the editing process.

Among the many items investigated by the Subcommittee were the role of subject matter specialists; hardware, software, and the data base environment; new technologies of data collection and editing, such as CATI and CAPI; current research efforts in the various agencies; and some recently developed editing systems such as at the Census Bureau and Statistics Canada.

In fulfilling its mission the Subcommittee followed a number of paths, including developing a questionnaire on survey editing practices, assembling several case studies of editing practices, investigating alternative editing systems and software, exploring research needs and practices, and compiling an annotated bibliography of literature on editing. The result of the Subcommittee's work is its report (1990), organized into 5 main chapters with several supporting appendices as follows:

Chapters	Appendices
I. Executive Summary	A. Questionnaire Responses
II. Background	B. Case Studies
III. Current Editing Practices	C. Software Functions Checklist
IV. Editing Software	D. Annotated Bibliography
V. Research on Editing	E. Glossary of Terms

After discussing some general topics pertaining to editing and to the Subcommittee's work, this paper summarizes some of the main results of a questionnaire on Current Editing Practices, designed, administered and compiled by the Subcommittee. The two papers immediately following address, respectively, the subjects of software developments and recent research findings in editing.

2. Data Editing -- Definition and Concepts

The subcommittee first addressed the definition of data editing. While no universal definition of survey data editing exists, the following working definition was developed:

Procedures designed and used for detecting erroneous and/or questionable survey data, with the goal of correcting (manually or electronically) as much of the erroneous data as possible (not necessarily all of the questioned data), usually prior to data imputation and summary procedures.

Thus data editing can be seen as a data quality improvement tool by which erroneous or highly suspect data are found and (if necessary) corrected. We have focused primarily on editing rather than imputation in our work, though in practice the boundary between these is not absolute.

3. Current Editing Practices

To obtain a profile of current editing practices in the various Federal statistical agencies, the subcommittee developed an editing questionnaire, which was completed for 117 Federal censuses and surveys representing 14 different Federal agencies. These 117 surveys were selected by subcommittee members, and thus they were not a scientific sample of all Federal surveys; however the Subcommittee felt that the 117 surveys represented a broad coverage of agencies and types of surveys or censuses that would present different editing situations.

The Subcommittee members primarily involved with the questionnaire and editing profile were Charles Day, Yahia Ahmed, George Hanuschak, Rita Hohenbrink and Renee Miller.

The questionnaire that was designed was a six-page document containing general questions about the particular survey as well as specific questions on editing. The report contains a complete listing of the questions asked, along with a tally of the results obtained for the 117 surveys, and should serve as a useful reference for the current (1990) state of data editing practice. A few of the major results follow.

Regarding general characteristics of the surveys, about three-fourths of the surveys are actually sample surveys, and the remaining one-fourth censuses. A wide range of frequencies of collection are represented, from daily to quinquennial. About one-fourth are completed by individuals, and three-fourths by establishments. While traditional means of data collection such as mail, personal and telephone interviews were most common, a small

proportion of the surveys used CATI, and some were administrative records.

Turning to editing, while the idea that there's no such thing as a free lunch seems to be as true of data editing as it is of anything else, there was wide variation in the actual cost of editing as a percent of total survey cost. The median editing cost for the surveys was more than one-third of the total cost of the survey. One of the interesting findings was that surveys of individuals had lower relative editing costs than surveys of establishments.

The questionnaire also elicited information on when in the survey process the editing occurs. For about two-thirds of the 117 surveys, most of the data editing takes place after data entry. Editing at the time of data entry is on the increase but not yet common.

Subject matter analysts play a large and important role in data editing. In about three-fourths of the surveys, subject matter analysts review all unusual or large cases. Only seven of the surveys had little or no intervention by subject matter specialists. In this regard, we found that surveys of establishments had heavier involvement from subject-matter specialists than surveys of individuals; and this could also be related to the finding, mentioned above, of lower editing costs in individual than in establishment surveys.

The degree of automation in data editing varies considerably among the surveys in our study. In about three-fifths of the surveys, automated edit checking is done, but error correction is performed by clerks or analysts. In about one-third of the cases, only unusual situations are referred to analysts. Only 3% of the surveys were totally automated, though all but 1% had at least some automation.

There are different types of edits that are applied to surveys. Almost all the surveys in our study use validation editing, which detects inconsistent data within a record. About five-sixths also use macro editing, where aggregated data are examined. The majority of surveys use other types of edits as well, such as range edits, edits using historical data, ratio edits, some of which may overlap. Additional information is also utilized in editing many of the surveys, such as comparisons with other surveys, comparison to a value estimated by regression analysis, or the use of interquartile measures.

Satisfaction with the current editing system varied widely. About half the respondents were satisfied with their current editing systems, and another one-fourth felt only minor changes were needed. The remaining one-fourth thought major changes were desired, with 5% of those being in favor of a complete overhaul.

Among those desiring improvements, those most frequently mentioned were:

- an on-line system for data editing,
- the use of prior periods' data to test the current period,
- more statistical edits,
- more sophisticated validation and macro editing,
- an audit trail,
- more automation, particularly automated error correction,
- user-friendlier systems,
- incorporation of imputation into the error package,
- evaluation of effects of data editing,
- reduction of the number of edit flags to follow up,
- incorporation of information on auxiliary variables,
- greater use of Expert Systems, and
- multivariate editing.

An audit trail, or a complete record of the original and corrected data, the edits failed and any other relevant information, is very helpful in monitoring and improving the editing process. The importance of an evaluation of the effects of editing on the data, and our current lack of knowledge of such effects, have also been noted by Bailar (1990).

4. Case Studies

In addition to the breadth of valuable information obtained from the questionnaire, the Subcommittee also felt that an examination of a relatively few surveys in greater depth would shed light on the complexity of the different editing situations in operation. Therefore several case studies are described, some in two-paragraph summary format and others in greater detail. These comprise Appendix B of the report. Anne Hafner and Yahia Ahmed had primary responsibility for preparation of the Case Studies.

5. Recommendations

The report lists a number of recommendations for future data editing practice, some general and some specific. Many of them fall into the following general categories.

The quality of an agency's existing editing practices and technology should be examined in the light of possible improvements or alternatives, with respect to such criteria as cost efficiency, timeliness, statistical defensibility, and accuracy.

Important recent developments in data processing, such as new microcomputers, workstations, local area networks, data base software, and mainframe linkages, should be

examined for their possible incorporation into the survey editing process.

Agencies should stay in communication with each other and with other professionals regarding their research in editing, particularly the development and implementation of new editing procedures and related methodologies such as data base technologies and expert systems.

References

Bailar, Barbara (1990), "Discussion of 'Survey Quality Profiles'", Seminar on the Quality of Federal Data, May 22, 1990, COPAFS. This Proceedings.

Groves, Robert (1990), "Towards Quality in a Working Paper Series on Quality", Keynote Address, Seminar on the Quality of Federal Data, May 22, 1990, COPAFS. This Proceedings.

Hanuschak, George, Yahia Ahmed, Laura Bauer, Charles Day, Maria Gon-zalez, Brian Greenberg, Anne Hafner, Gerry Hendershot, Rita Hohenbrink, Renee Miller, Tom Petkunas, David Pierce, Mark Pierzchala, Marybeth Tschetter and Paula Weir (1990), Data Editing in Federal Statistical Agencies, Statistical Policy Working Paper 18, Statistical Policy Office, Office of Management and Budget, Washington, DC.

EDITING SOFTWARE
(An excerpt from Chapter IV of Working Paper 18)

Mark Pierzchala
National Agricultural Statistics Service

A. Introduction

For most surveys, large parts of the editing process are carried out through the use of computer systems. The task of the Software Subgroup has been to investigate software that in some way incorporates new methodologies, has new ways of presenting data, operates in recently developed hardware environments, or integrates editing with other functions. In order to fulfill this charge, the Subgroup has evaluated or been given demonstrations of new editing software. In addition, the Subgroup has developed an editing software evaluation checklist that appears in Appendix C of Statistical Policy Working Paper 18. This checklist contains possible functions and attributes of editing software, which would be useful for an organization to use when evaluating editing software.

Extremely technical jargon can be associated with new editing systems; and new approaches to editing may not be familiar to the reader. The purpose of section B is to explain these approaches and their associated terminology as well as to discuss briefly the role of editing in assuring data quality.

A distinction must be made between generalized systems and software meant for one or a few surveys. The former is meant to be used for a variety of surveys. Usually there is an institutional commitment to spend staff time and money over several years to develop the system. It is hoped that the investment will be more than recaptured after the system is developed through the reduction in resources spent on editing itself and in the elimination of duplication of effort in preparing editing programs. Some software programs have been developed that address specific problems in a particular survey. While the ideas inherent in this software may be of general interest, it may not be possible to apply the software directly to other surveys. Section C of Chapter IV of Working Paper 18 describes three generalized systems in some detail, and then briefly describes other systems and software. These three systems have been used or evaluated by Subgroup members in their own surveys.

New and exciting statistical methodology is also improving the editing process. This includes developments in detecting outliers, aggregate level data editing, imputation strategy, and statistical quality control of the process itself. The implementation of these activities, however, requires that the techniques be encoded into a computer program or system.

B. Software Improving Quality and Productivity

Reasons for the Development of New Editing Software

Traditional editing systems do not fully utilize the talents or expertise of subject matter specialists. Much of their time may be spent in dealing with unimportant or spurious error signals and in coping with system shortcomings. As a result, the specialist has less time to deal with important problems. In addition, editing systems may be able to give feedback on the survey itself. For example, a pattern of edit failures may suggest misunderstandings by the respondent or interviewer. If this is recognized, then the expertise of the specialist may then be used to improve the survey itself.

Labor costs are a large part of the editing costs and are either steady or increasing, whereas the cost of computing is decreasing. In order to justify the heavy reliance on people in editing, their productivity will have to be improved through the use of more powerful tools. However, even if productivity is improved, different people may do different things in similar situations. If so, this makes the process less repeatable (reproducible) and more subject to criticism. When work is done on paper, it is hard to track, and it is impossible to estimate the effect of editing actions on estimates. Finally, some tasks are beyond the capability of human editors. For example, it may be impossible for a person to maintain the multivariate frequency structure of the data when making changes.

These reasons and several others are commonly given as explanations for the increased use of computer software to improve the editing process. It is in the reconciliation of these two goals, (the increased use of computers for some tasks and the more intelligent use of human expertise), that the major challenge in software development lies. There will always be a role for people, but it will be modified. One positive feature of new editing software is that it can often improve the quality of the editing process and productivity at the same time.

Ways That Productivity Can Be Improved

One way to improve productivity is to break the constraints imposed by computer systems themselves. The use of mainframe systems for editing data is widespread. In some cases, however, an editor may not use the system directly. For example, error signals may be presented on paper printouts, and changes entered by data typists. Processing costs may dictate that editing jobs are run at low priority, overnight, or even less frequently. The effect of the changes made by the editor may not be immediately

known: thus, paper forms may be filed, taken from files, and re-filed several times.

The proliferation of microcomputers promises to eliminate many of these bottlenecks, while at the same time it creates some challenges in the process. The editor will have direct access to the computer, and will be able to prioritize its use. Once the microcomputer is acquired, user fees are eliminated; thus resource-intensive programs such as interactive editing can be employed, provided the microcomputers are fast enough. Moving from a centralized environment (i. e., the mainframe) to a decentralized environment (i. e., microcomputers) will present challenges of control and consistency. In processing a large survey on two or more microcomputers, communications will be necessary. This will best be done by connecting them into a Local Area Network (LAN).

New systems may reduce or eliminate some editing tasks. For example, where data are edited in batch and error signals are presented on printouts, a manual edit of the questionnaires before the machine edit may be a practical necessity. Editing data and error messages on a printout can be a hard, unsatisfactory chore because of the volume of paper and the static and sometimes incomplete presentation of data. The purpose of the manual edit in this situation is to reduce the number of machine-generated error signals. In an interactive environment, information can be efficiently presented and immediately processed. The penalty associated with machine-generated signals is greatly reduced. As a result, the preliminary manual edit may be eliminated. In addition, questionnaires are handled only once, further reducing filing and data entry tasks.

Productivity may be increased by reducing the need for editing after data are collected. Instruments for Computer Assisted Telephone Interviewing (CATI), Computer Assisted Personal Interviewing (CAPI), and on-site data entry and editing programs are gaining wider use. Routing instructions are automatically followed, and other edit failures are verified at the time of the interview. There may still be many error signals from suspicious edits, however, the analyst has more confidence in the data and is more likely to let them pass.

There are two major ways that productivity can be improved in the programming of the editing instruments. First is to provide a system that will handle all, or an important class, of the agency's editing needs. In this way the applications programmer need not worry about systems details. For example, in an interactive system, the programmer does not have to worry about how and where to flag edit failures as it is already provided. The programmer only codes the edit specification itself. In addition, the end-user has to learn only one system when editing different surveys. Second is the elimination of multiple specification and programming of variables and edits. For example, if data are

collected by CATI, and edited with another system, then essentially the same edits will be programmed twice, possibly by two sets of people. If the system integrates several functions, e. g., data entry, data editing, and computer assisted data collection, then one program may be able to handle all of these tasks. This integration would also reduce time spent on data conversion from one system to another.

Systems That Take Editing and Imputation Actions

Some edit and imputation systems take actions usually reserved for people. They choose fields to be changed and then change them. The human element is not removed, rather this expertise is incorporated into the system. One way to incorporate expertise is to use the edits themselves to define a feasible region. This is the approach outlined in a famous article by Fellegi and Holt (1976). Edits that are explicitly written are used to generate implied edits. For example, if $100 < x / y < 200$, and $3 < y / z < 4$, are explicit edits, then an implied edit obtained algebraically is $300 < x / z < 800$. Once all implied edits are generated, the set of complete edits is defined as the union of the explicit and implied edits. This complete set of edits is then used to determine a set of fields to be changed for every possible edit failure. This is called error localization. An essential aspect to this method is that changes are made to as few fields as possible, or alternatively, to the least reliable set of fields which are determined by weights given to each field.

The analyst is given an opportunity to evaluate the explicit edits. This is done through the inspection of the implied edits and extremal records (the most extreme records that can pass through the edits without causing an edit failure). In inspecting the implied edits, it may be determined if the data are being constrained in an unintended way. In inspecting extremal records, the analyst is presented with combinations of the most extreme values possible that can pass the edits. The human editor has several ways to inject expertise into this kind of a system: (1) the specification of the edits; (2) the inspection of implied edits and extremal records and then the re-specification of edits; (3) the weighting of variables according to their relative reliability.

There are some constraints in systems that allow the computer to take editing actions. Fellegi and Holt systems cannot handle certain kinds of edits, notably nonlinear and conditional edits. Also algorithms that can handle categorical data cannot handle continuous data and vice versa. Within these constraints (and others), most edits can be handled. For surveys with continuous data, a considerable amount of human attention may still be necessary, either before the system is applied to data or after.

Another way that computers can take editing actions is by modeling human behavior. This is the "expert system" approach. For example, if typically maize yields average 100 bushels per acre, and the value 1,000 is entered, then the most likely correction is to assume that an extra zero was typed. The computer can be programmed to substitute 100 for 1,000 directly and then to re-edit the data.

Ways That Data Quality Can Be Improved or Maintained

It is not clear that editing done after data collection can always improve the quality of data by reducing non-sampling errors. An organization may not have the time or budget to recontact many of the respondents or may refrain from recontacts in order to reduce respondent burden. Additionally, there may be cognitive errors or systematic errors that an edit system cannot detect. Often, all that can be done is to maintain the quality of the data as they are collected. To use the maize yield example again, if the edit program detects 1,000 bushels per acre, and sets the value to 100 bushels per acre, then the edit program has only prevented the data from getting worse. Suppose the true value was really 103 bushels per acre. The edit and imputation program could not get the value closer to the truth in this case. Detecting outliers is usually not the only problem. The proper action to take after detection is the more difficult problem. One of the main reasons that Computer Assisted Data Collection is employed is that data are corrected at the time of collection.

There are a few ways that an editing system may be able to improve data quality. A system that captures raw data, keeps track of changes, and provides well conceived reports, may provide feedback on the performance of the survey. This information can be used to improve the survey in the future. To take another agricultural example, farmers often harvest corn for silage (the whole plant is harvested, chopped into small pieces, and blown into a silo). Production of silage is requested in tons. Farmers often do not know their silage production in tons. Instead, the farmer will give the size (diameter and height) of all silos containing silage. In the office, silo sizes are converted into tons of production. If this conversion takes place before data are entered, then there is no indication from the machine edit of the extent of this reporting problem.

Another way that editing software can improve the quality of the data is to reduce the opportunity cost of editing. The time spent on editing leaves less time for other tasks, such as persuading people to participate, checking overlap of respondents between multiple frames, and research on cognitive errors.

Ways That Quality of the Editing Process Can Be Defended or Confirmed

There is a difference between data quality and the quality of the editing process itself. To refer once again to the maize yield example, a good quality process will have detected the transcription error. A poor quality process might have let it pass. Although neither process will have improved data quality, the good quality process would have prevented their deterioration from the transcription error. Editing and imputation have the potential to distort data as well as to maintain their quality. This distortion may affect the levels of estimates and the univariate and multivariate distributions. A high quality process will attempt to minimize distortions. For example, in Fellegi and Holt systems, changes to the data will be made to the fewest fields possible and in a way such that distributions are maintained.

A survey organization should be able to show that the editing process is not abusing the data. For editing after data collection, this may be done by capturing raw (unedited) data and keeping track of changes and the reasons for change. This is called an audit trail. Given this record keeping, it will be possible to estimate the impact of editing and imputation on expansions and on distributions. It will also be possible to determine the editor effect on the estimates. In traditional batch mode editing on paper printouts, it is not unusual for two or more specialists to edit the same record. For example, one may edit the questionnaire before data entry while another may edit the record after the machine edit. In this case, it is impossible to assign responsibility for an editing action. In an on-line mode one person handles a record until it is done. Thus all changes can be traced to a person. For editing at the time of data collection, (e. g., in CATI), it may be necessary to conduct an experiment to see if either the mode of collection, or the edits employed, will lead to changes in the data.

A high quality editing process will have other features as well. For example, the process should be repeatable, in time and in space. This means that the same data passed through the same process in two different locations, or twice in one location, will look (nearly) the same. The process will have recognizable criteria for determining when editing is done. It will detect real errors without generating too many spurious error signals. The system should be easy to program in and have an easy user interface. It should promote the integration of survey functions such as micro- and macro-editing. Changes made by people should be on-line (interactive) and traceable. Database connections will allow for quick and easy access to historical and sampling frame data. An editing system should be able to take actions of minor impact without human intervention. It should be able to accommodate new advances in statistical editing methodology.

Finally, quality can be promoted by providing statistically defensible methods and software modules to the user.

Acknowledgements

Other members of the Editing Software Working Group for Working Paper 18 were Tom Petkunas, Bureau of the Census, Gerry Hendershot, National Center for Health Statistics, Charles Day, Internal Revenue Service, Marybeth Tschetter, Bureau of Labor Statistics, and Rita Hohenbrink, National Agricultural Statistics Service.

RESEARCH ON EDITING

Yahia Ahmed
Internal Revenue Service

Introduction

This paper is one of three papers presented in a session organized to present topics from the Statistical Policy Working Paper 18, "Data Editing in Federal Statistical Agencies." The Subcommittee on Data Editing in Federal Statistical Agencies was established by the Federal Committee on Statistical Methodology to document, profile and discuss data editing practices in Federal surveys. To effectively accomplish its mission, the subcommittee was divided into four major groups: Editing Profile, Case Studies, Editing Software, and Editing Research.

The purpose of this paper is to present briefly the goals, findings and recommendations of the Editing Research Group. A more detailed description of editing research is provided in Chapter V of the Working Paper.

The goals of the Editing Research Group were to identify areas in which improvements to edit systems would prove most useful, to describe recent and current research activities designed to enhance edit capabilities, to make recommendation for future research and to develop an annotated bibliography on editing.

Areas Which Need Improvement

The Editing Research Group used two sources of information to identify areas which need improvement. The first source was the editing profile questionnaire which was administered to managers of 117 Federal surveys covering 14 different agencies. This questionnaire included questions about edit improvements. One question asked was "For future applications, what would you like your edit system to do that it doesn't do now?" The second source was discussions with those responsible for edit tasks within a number of Federal agencies. The following areas emerged as priorities:

- o More on-line edit capabilities
- o Better ways to detect potentially erroneous responses
- o More sophisticated and extensive macro-editing
- o Evaluation of the effect of data editing.

Areas of Edit Research

Much editing research has been conducted in national statistical offices around the world. It is these organizations, which conduct huge and complicated surveys, that have the most to be gained from developing new systems and techniques. They also have the resources upon which to draw for this development.

One area of current research interest is that of "on-line" edit capabilities". BLAISE, SPEER, and PEDRO discussed in the preceding paper are examples of such research activities.

A second area of active research is in the detection of potentially erroneous responses. The method most commonly used is to employ explicit edit rules. For example, edit rules may require that:

- 1) The ratio of two fields lie between prescribed bounds,
- 2) various linear inequalities and/or equalities hold, or
- 3) the current response be within some range of a predicted value based on a time series or other models.

Edit rules and parameters are highly survey specific. A related area of editing research is the design of edit rules and the development of methods for obtaining sensitive parameters.

In order to make sure that all errors are flagged, often many unimportant error flags are generated. These extra flags not only take time to examine but also distract the reviewer from important problems. These extra flags are generated because of the way that the error limits are set. A related area of research focuses on developing statistical editing techniques to reduce the number of error flags, while at the same time, ensuring that not many errors escape detection. Several research studies in which different statistical techniques (such as clustering, exponential smoothing and Tukey's biweight) to detect potentially erroneous responses or to set error bounds are described in the working paper.

In contrast to the rule-driven method for the detection of potentially erroneous response combinations within a record, one alternative procedure is to analyze the distribution of questionnaire response. Records which do not conform to the observed distribution are then targeted as outliers and are selected for review. Although there has been research interest in this method, no application of these multivariate methods was found.

Recommendations

The most important recommendation is that agencies recognize the value of editing research and place in high priority on devoting resources to their own research, to monitoring developments in data editing at other agencies and elsewhere and to implement improvements.

Often innovations in editing methods made by survey staff are viewed as enhancements to processing for that particular survey and little thought is given to the broader applicability of methods developed. Accordingly, survey staff do not prepare discussion of new methods for publication. We encourage survey staff to take the time to describe their work and publish them in order to share their experiences with others who may be working under similar conditions. It is often in such articles that methods which may be applicable to more than one survey are first introduced and described.

The survey on editing practices indicated that there was little analysis of the effect of editing on the estimates that were produced. Considering that the cost of editing is significant for most surveys, this is clearly an area in which more work is required. A related issue is the need to attempt to determine when to edit and not to edit.

Clearly, all the errors are not going to be found and we should not attempt to find them all. Therefore, there is a need to design guidelines for determining what is an acceptable level of editing.

Another neglected research area in this country concerns the editing of data at the time they are keyed from mail responses. This area is usually discussed in the setting of quality control; however, it is an area that can benefit from further research from the perspective of data editing.

Annotated Bibliography

It is quite difficult to provide a complete assessment of current research activities in the area of editing because so much of the research, progress, and innovations are described only in specific documentation. However the group was able to identify 86 references which describe research efforts over the past years. Appendix D of the working paper contains the annotated bibliography. The annotations are brief and are only intended to give a very general idea of the paper's content. The appendix provides a valuable source of information on the editing literature. In addition, it includes papers which describe the underlying methods, the software, proposed uses, and possible

advantages of three generalized editing software systems -- GEIS, BLAISE and SPEER.

Acknowledgements

Other members of the Editing Research Group for Working Paper 18 were Laura Bauer, Federal Reserve Board, Brian Greenberg, Bureau of the Census, Renee Miller, Energy Information Administration, David Pierce, Federal Reserve Board, Paula Weir, Energy Information Administration.

DISCUSSION

Charles E. Caudill
National Agricultural Statistics Service

As Administrator of a Federal-State Cooperative Statistical Agency, I am quite impressed with the information contained in OMB Statistical Policy Working Paper No. 18 on Data Editing in Federal Statistical Agencies. The working paper thoroughly documents many existing editing practices, generalized editing software developments and provides a detailed software evaluation protocol. In addition, it covers current research activities on editing, provides an annotated bibliography and has a good executive summary including recommendations.

I believe that this report, if read and seriously considered by federal survey managers and administrators, can have a substantial effect on improving productivity. Thus, "precious" resources could be freed up to more formally address nonsampling errors, quality control, and total survey error models, measurements and structures. In my opinion, if there was ever a report that survey administrators should take seriously, this is it.

There are several more detailed comments and observations that I have about working paper number 18. The data on the costs of editing was intriguing. My observation is that there may be an upward bias in the data and some non-editing cost may have been included. However, even if this is the case, there obviously is still plenty of room for productivity gains in the editing process. With the proliferation of personal computer networks and data base software, there is substantial potential to improve the productivity of editing systems by being on-line and providing the editor with immediate screen feedback and re-editing of their proposed changes.

Recent computer processing technology advances also make the use of audit trails more available for more users. Inexpensive audit trails provide the capability to analyze and conduct research on the effects of editing on the estimators and also on the overall performance of the survey as well.

The detailed checklist of edit software system features in Appendix C of working paper 18 will be beneficial to both the development of new systems and maintenance and evaluation of existing systems. The annotated bibliography of articles and papers on editing presented in Appendix D will be valuable for researchers and system developers as a substantial source of literature and information.

Working paper 18 certainly demonstrated that current data editing practices are labor intensive. Many remain mainframe and batch oriented, with multiple passes of the data. Also, I think that there may be a tendency to stay with existing systems too long.

My final comments are on total quality management of surveys. As an Administrator, one of my major concerns is with the quality of the final products and reports that the Agency delivers to the public. Thus, if the editing process can be made more efficient, without degrading accuracy, then that adds to the potential of using the saved resources on other important areas of the survey process. Total quality management techniques applied to surveys are useful tools in efficiently identifying the most important potential sources of survey error.

DISCUSSION

Richard Bolstein
George Mason University

The serious impact that erroneous survey data can have on results, the fact that the number of errors tend to increase with the size and complexity of the survey, and the relatively large proportion of survey costs currently required to edit and correct data, make the need for new and improved methods of data editing imperative. To this end, the authors have done a laudable job in researching methods currently used, presenting several case studies, testing and discussing the advantages and disadvantages of some current and developing editing software, and providing a synopsis of current research.

A working definition of editing was clearly necessary in this study, since, among other things, in order to estimate costs of editing, a fairly rigorous definition of the scope of editing was required. The working definition used by the authors, namely, "procedure(s) designed and used for detecting erroneous and/or questionable survey data with the goal of correcting as much of the erroneous data as possible, usually prior to data imputation and summary procedures" is quite suitable for this purpose. We should keep in mind, however, that while it feels comfortable to clean up erroneous data prior to imputation for missing data, in practice the two are often intertwined.

The paper states that the cost of editing was available for 40% of the 117 surveys in the sample, and cost estimates were possible for an additional 40%. It was reported that between 75% and 80% of these surveys had editing costs of at least 20% of total costs. It is not too meaningful to compare the relative costs of editing across all types of surveys, however, since one would naturally expect that these costs would be higher in less expensive surveys (such as mail or administrative records) than in expensive surveys (such as personal interview, surveys of institutions), as found by the authors. Thus, it would be more informative if the relative cost figures cited above were reported by survey type. Another factor that can account for a large percentage of editing costs is the presence of a relatively large number of questions requiring open-ended responses and subsequent coding of the responses. But although the distribution of the relative cost of editing may vary considerably, there is no doubt that editing is costly and methods to reduce this cost and improve data quality are much needed.

Finally, no discussion of the costs of editing is complete without determining what percentage is due to bad data that should not have occurred but for inadequate interviewer training, poor supervision and quality control of interviewers, and simple common

sense errors. For these are errors which should not have occurred and should be deducted from the cost of editing in the estimates of the surveys above, since they are likely to have varied considerably.

Although elimination of such unnecessary errors was not part of the project of the three authors, it seems appropriate in a discussion of improving data editing procedures to mention ways in which the need for editing can be reduced. To illustrate an example of a common sense error that should be eliminated, in a certain survey, the sponsor of which I will not name, fishermen are interviewed and their catch is weighed and measured. The interviewer is supposed to record weight in kilograms, but the scale used shows weight in both pounds and kilograms. As expected, frequent errors occur. The obvious solution is to use a scale that only shows kilograms, but when I suggested this to the survey firm, the response was "no one makes such a scale". When I then suggested taping over the side of the scale showing pounds, the reply was "but the fishermen want to know what their fish weigh in English". Finally, I suggested taping over the kilogram side of the scale, have the interviewer record the weight in pounds, and have the data entry program convert it to kilograms. The response to this suggestion I am sure you have all heard before: "well, that's the way we're used to doing it". There are numerous other examples of course (for example, in some surveys interviewers are required to record the hour in military time).

The most promising methods to reduce editing costs and improve data quality (after elimination of the unnecessary errors) are found in interactive data entry software and in general editing software systems. These methods seem appropriate for large, complex surveys, or surveys which are repeated. For small one-time surveys the cost of purchasing, learning, and programming the software will most likely outweigh the savings, as this is even true with CATI. But this is generally not the case with surveys gathering Federal Data. The three generalized editing software systems studied in detail by Mark Pierzchala seem very promising, especially BLAISE because of its generality and ability to handle both categorical and continuous data. GEIS and SPEER are specific to economic type surveys.

To what extent can graphics or other theoretical tools be used in editing systems? The STAR WARS software described uses graphics to compare edited values with the originals, but not to detect outliers. The parallel coordinate system for graphic displays of high-dimensional data [see Miller and Wegman (1989), Wegman (1990)] may be used to detect outliers. Yahia Ahmed noted that analysis of the multivariate distribution of questionnaire responses to flag records that don't conform to the distribution as outliers has been infrequently used, no doubt due to its complexity. I believe that graphical methods for detecting outliers will meet with more acceptance than the multivariate analysis approach has but it would

not be cheap (time-wise) and probably would be best used as a final check rather than at the front-end of the editing task.

Finally, I have two recommendations. In view of the increasing abundance of software we will see in the future, we should construct a standard set of test data sets for evaluating present and future software editing systems. Secondly, a one or two-day demonstration seminar of some of these systems would be well received.

References

Miller, J.J. and Wegman, E.J. (1989), "Construction of line densities for parallel coordinate plots", Technical Report No. 53, Center for Computational Statistics, George Mason University.

Wegman, E.J. (1990), "Hyperdimensional data analysis using parallel coordinates", Journal of the American Statistical Association, to appear.

Session 6

COMPUTER ASSISTED STATISTICAL SURVEYS

OVERVIEW OF COMPUTER ASSISTED SURVEY INFORMATION COLLECTION

Richard L. Clayton
U. S. Bureau of Labor Statistics

This section provides a summary of Working Paper 19 on Computer Assisted Survey Information Collection (CASIC). For additional information, we encourage you to see this document.

The power of rapid calculating has been applied to virtually every phase of the survey process, including sample design and selection, and estimation. The most important implication of these applications is that survey practitioners are allowed to consider a growing range of techniques which were not affordable prior to the availability of inexpensive and fast calculating capability.

The field of computer assisted collection applications may be the area of greatest and most rapid change in survey methods. This field includes the rapidly expanding variety of applications based on the availability of powerful and inexpensive computers. Most familiar of the new techniques are CATI and CAPI. However, a variety of other collection methods are being developed across the Federal government's statistical agencies, including Touchtone Data Entry, Prepared Data Entry and more recently, Voice Recognition Entry.

High quality published data begins with collecting high quality data from our respondents. Much of survey processing addresses, and compensates for, weaknesses in the quality of the collected data and the data we do not collect. Those methods which capture data quickly and accurately should be developed which allow respondents to answer our questions accurately and quickly. With this in mind, we provided the results of research and development activities using new technological features throughout the Federal government seeking new data collection methods, and in modifying the old, to improve the quality of data collection.

For the purposes of this report, we defined computer assisted survey information collection methods as those using computers as a major feature in the collection of data from respondents, and in transmitting of data to other sites for post-collection processing.

Goal: The overall goal of Working Paper 19 was to provide information on new data collection methods to challenge Federal survey managers to reconsider their operations in light of recent changes in survey methods available, or made attainable through changing technology; to reassess their methods of accomplishing the common goal of providing the critical information to the public which is accurate, timely and relevant. We hope that by sharing information and experiences, that others may gain and forward the overall effectiveness of governmental activities.

Objectives: The primary objective is to describe emerging methods of interactive electronic data collection, the potential benefits, and current examples of its use in Federal surveys. In describing current uses and tests, a secondary objective is to pose questions about the implications of use of computer assisted methods and try to suggest some answers. These questions involve such factors as quality, costs, and respondent reaction to computerized surveys.

Scope: The survey operations included in this report includes all of the activities and tasks from the transmittal of the questionnaire, conduct of the interview, data entry, editing and followup for nonresponse or edit reconciliation.

The last major survey operation to benefit from automation is data collection. Computers were first applied to collection using mainframes to control certain aspects of telephone collection, and Computer Assisted Telephone Interviewing (CATI) was born. The first applications of CATI stimulated new research worldwide evaluating the impact on of CATI on the survey error profile and costs. CATI is now used to assist interviewers in all collection activities, including scheduling calls, controlling detailed interview branching, editing and reconciliation, providing much greater control over the collection process and reducing many sources of error. At the same time, a tremendous amount of information is captured by the computer providing additional insight into the data collection process.

The ongoing advances in computer technology, and particularly the advent of microcomputers, continue to offer additional opportunities for improving the quality of published data. The first portable computers were quickly pressed into service to duplicate the advantages of CATI in a personal visit environment. Thus, Computer Assisted Personal Interviewing (CAPI) was launched from the work in CATI.

While CATI and CAPI represent advances for surveys requiring interviewers, microcomputers are now finding important roles in self-administered questionnaires, where interviewers are not needed.

Prepared Data Entry (PDE), developed by the Energy Information Administration, allows respondents which have a compatible microcomputer or terminal to access and complete the questionnaire directly on their screen.

Touchtone Data Entry (TDE), developed at the Bureau of Labor Statistics, allows respondents to call a toll-free telephone number. Questions posed by a computer are answered using the keypad of their touchtone telephone. The machine repeats the answers for verification with the respondent which are stored in a database. TDE systems are now commonplace for bank transfers, and

telephone call routing, as examples. We have just applied existing technology to the data collection process.

As an extension of this approach, techniques have been developed more recently allowing respondents to answer the questions by speaking directly into the telephone. The incoming sounds are matched to known patterns recognizing the digits and the words "yes" and "no". Voice Recognition Entry (VRE), as this is known, is not the distant future. The Bureau of Labor Statistics is currently conducting live tests where this method is being warmly received by respondents as natural and convenient.

Both TDE and VRE offer inexpensive data collection where the respondents initiate the calls, enter and verify the data. Refinements to procedures will now focus on minimizing nonresponse prompting activities.

Respondent Burden: For many respondents, the use of automated methods can actually reduce the collection burden placed on them. For example, use of Prepared Data Entry, where respondents interact with computer screens, provides a single set of step-by-step procedures with on-line editing to prevent inconsistent or incorrect reporting, thus reducing the need for expensive and troublesome recontacts. Also, these methods have, in some cases, substantially reduced the time taken to provide complex data for large establishments. Similar methods may be applied to other surveys covering large establishments where the one-time costs of data conversion to a standard format would be cost-effective, especially in repeated surveys.

Quality: Automated collection allows for improved control yielding reduced error from several sources including errors caused by the respondent, the interviewer, and post collection processes such as key entry error. The instant status capabilities of CATI, for example, provide stronger intervention features for nonresponse prompting, reducing nonresponse error.

In deciding which collection method to use, quality can become a relative concept that is affected by a tradeoff between cost and benefit. The choice of a data collection method is usually based on a combination of performance and cost factors determining affordable quality. For traditional collection methods, these factors and the decision-making process are fairly well known. Now, these new methods discussed in Working Paper 19 expand the array of potential collection tools and challenge the survey designer to reevaluate old cost/performance assumptions.

Costs: The data collection process is composed of a few major activities, including transmitting and receiving the questionnaire, data entry, editing and nonresponse prompting. The labor and nonlabor costs will vary depending on the method used. For example, under mail collection virtually each action is conducted

manually and postage is the dominant nonlabor cost. By contrast, CATI operations can minimize postage costs reduces many of the expensive mail handling operations. However, CATI adds new costs in the form of telephone line charges and computers (including systems design and ongoing maintenance). Self-response methods, such as TDE, VRE and PDE collection, reduce postage, the manual mail operations and the labor involved in CATI interview activities, but may still require edit reconciliation and nonresponse followup.

Thus, the factors of production, and the composition of each those inputs vary greatly among the existing and newer techniques. Many factors can change in a short period. Only a few years ago, automation costs were driven by the scarcity of mainframe hardware capacity. Now, the costs of automation are driven by the labor involved in developing specialized systems dominates automation costs. Portable and desktop microcomputers were not widely available at the beginning on this decade. Now, microcomputers are widely available, very inexpensive and extremely powerful.

Old assumptions about costs need to be reevaluated. Labor and postage costs have risen steadily in recent years, while capital costs, such as microcomputers and telephone services have been declining.

The decision on which collection mode to use, or which combination, will depend on the particular survey application and the existing cost structure. However, it is important to view such investments over the long-term as the relative costs of each of the inputs do not remain constant over time. Survey managers should periodically review old assumptions in light of new technology and project operating costs over the reasonable foreseeable future in deciding not to investigate new methods.

Users: Automated data collection includes three major groups of people: the respondents, the interviewers and the designers and developers of the system and procedures for collection. This report covers the essential factors involved in successfully including the requirements of each group.

Respondents: The respondent must be considered the primary user of any survey vehicle, whether automated or not, and all aspects of the response environment must be developed with the respondent in mind. The cooperation of the respondent is the single most critical factor in survey operations. Respondents must be treated with the greatest care. We must consider our respondents as a Customer, after all, if our survey vehicle doesn't "sell", if the questionnaire is not successful in getting an accurate response, we will have no input for the rest of our production process.

Even one-time surveys must strive to leave the respondent with the feeling of contribution and importance, and most of all, a willingness to participate in other surveys in the future if called on. Thus, our primary job is to develop techniques which allow the respondent to complete the survey completely and accurately and with a minimum level of burden.

The use of these collection methods, while bringing improvements in the quality of collected data, has entailed other challenges. These automated collection methods are made possible through the close interaction of subject matter experts, statisticians, and computer scientists. To effectively use these methods, each of these groups learned the basic tenants of the others. This close relationship will only continue to grow, with advances in each field aiding advances in the others.

Interviewers: The second most important user is the interviewer. The systems provided to assist in the interview process must be easy to use, must work infallibly and must actually provide improvements in his or her work environment. Interviewers must feel as they are the most valuable feature in the interview, that the machine is merely a tool to expedite and simplify their work. This is not always an easy task.

Survey practitioners: We are the third major group of users. The decisions made early in the development process will carry over into the ongoing use and maintenance of the system.

Systems designers face difficult choices, such as building customized systems from scratch versus linking standardized "off the shelf" routines or commercial packages. The inevitable limitations would have to be traded off against reduced maintenance and lower start up costs.

Automated collection methods can also improve data quality. All of the methods discussed could be designed to include on-line editing to prevent impossible and inconsistent entries. Some of these methods, such as TDE and VR, improve data quality by verifying recorded data with the respondent.

These are potential improvements. The final impact of quality lies in the up front planning and execution. This place responsibility for clearly defining and controlling the collection environment directly with the survey designer.

Future: The future application of these techniques is limited only by our creativity and initiative of program managers and planners. The "case studies" serve to illustrate the options available, and will surely raise many more questions for further investigation.

We hope that the discussion of technological advances generates discussion and stimulates creative, new applications to the whole range of governmental information collection activities.

In addition to the methods described here there are other advances in technology which hold potential for vastly changing data collection. Integrated Services Digital Network (ISDN) is a powerful network system which will provide simultaneous transmission of sound, video and data. The result could be a change in the way some surveys are conducted offering all of the benefits of personal interviewing with the lower costs of telephone interviewing.

You have heard a several different collection methods described and discussed which are currently available. And you can see that the pace of change will accelerate and match changes in technology. So what does the future hold?

You have to ask yourself how your survey operations will be conducted in 5 or perhaps 10 years. In doing so, ask yourself how things were done 5 or 10 years ago. What sorts of things have happened and what were their implications?

A COMPARISON BETWEEN CATI AND CAPI

Martin Baum
National Center for Health Statistics

Introduction

I will describe for you some of the critical factors one must consider when deciding whether to conduct a survey by either CATI or CAPI. I also will try to indicate the similarities and differences between these two methods of survey data collection automation.

Definition

Let me first define each of the methods. Computer Assisted Telephone Interviewing (CATI) is a computer assisted survey process which uses the telephone for voice communications between the interviewer and the respondent. Computer Assisted Personal Interviewing (CAPI) is a personal interview usually conducted at the home or business of the respondent using a portable computer.

Rationale

The rationale for the development and for your use of these methods are based primarily on reasons of improved data quality and improved timeliness of data release. Cost is a factor, but in our experience, it has been a break-even situation; the cost of automating has equaled the savings. This result has been due primarily to the high cost of software development.

Factors

The following are critical factors that must be considered in addition to those of improved data quality and timeliness, and cost when deciding whether to use CATI or CAPI for your survey data collection. I will discuss each of these factors in some detail.

Hardware CATI

Initially CATI was developed as a mainframe application, but as computer technology changed, CATI moved to the mini computer and then to a networked micro computer application. The investment in hardware has steadily decreased without any loss of capability. Telephone technology, which impacts telephone availability is important to the CATI application - no phone no respondent.

Hardware CAPI

The most important computer hardware criteria for a CAPI application are generally quite different from those that would be critical to most other applications. The major reason is the role that environmental conditions play in the selection of CAPI hardware. The fact that CAPI is a personal interview situation, usually taking place in or at the home of the respondent, dictates a number of possible circumstances under which the interview will be conducted.

For example, screen visibility becomes a paramount criterion because of the environmental conditions. Interviews will take place under all types of lighting conditions; outside in bright sunlight, twilight, and normal light, and inside under lamp light, fluoresce light, and bear bulb.

Weight is especially critical because of the variety of environmental conditions. Interviewers may be conducting the survey in an urban setting where the computer will be carried up and down the stairs of apartment houses; or in a suburban setting where the computer is carried many blocks; or in a rural setting where the computer is carried long distances from car to house. In any of these conditions, the computer is moved in and out of a car many times. This situation is further compounded by the fact that the interviewer must also carry considerable paper e.g. back-up paper questionnaires in case the computer fails, letters of explanation, introduction, and thank you. Carrying all of this weight in and out of cars and up and down steps all day is no easy job, particularly if the computer and back up battery weighs 10 plus lbs. and the paper weighs an additional 5 lbs or more.

For a household type survey, the interviewers are generally reluctant to ask for the respondent's permission to use power for the computer because of fear of possibly losing the interview. Also, surveys frequently are conducted outside of the house where no power is available. Many of our surveys can last as long as 2-4 hours. Consequently, battery life is critical.

Environmental conditions often impact the ergonomics of the hardware. Consider a survey interview conducted where the computer must be placed on the interviewer's lap. This situation would be quite difficult if the computer were either top heavy when open or the interviewer was small and the computer's depth long. Balancing would be a problem. Also consider the door step interview with a 10 lb. clam shell design computer.

Software

Now let's discuss the most costly factor in the CATI/CAPI decision - software. There are four components to the CATI/CAPI

software: Questionnaire, Case Management, Output Reporting, and Authoring System.

The questionnaire component refers to the software that places each question in the survey on the computer screen in the proper sequence with the appropriate information (i.e. prompts) and allows the entry of an answer or answers to the question with edits on those answers such as; range, specific values, consistency with another question's answer. This software should also contain on-screen help and if necessary, rostering.

The case management component is the software that allows the interviewer to keep track of the status of the survey interview; that is, is the interview complete?; if the interview is not complete, what has been completed and what is the next question to be asked?; is the interview a partial interview or is the interview to be completed later?; what sections of the survey are mandatory?; and in some instances, interviewer assignments. In the case of CATI, case management software also would provide the sample selection and dialing of the phone number.

The output reporting component is often either overlooked or given minimal consideration. This is a big mistake. Collection of the data is not very useful if the data cannot be easily accessed for analysis. Output reports can be categorized as either survey questionnaire statistics or management statistics. The level of detail and complexity can vary significantly. Survey questionnaire reporting can be as little as the ability to place the data into specific analysis software file format e.g. SAS or can include actual analyses.

Management statistics can be extremely useful for the conduct of the survey data collection. For example, data can be automatically collected on the time to complete a section of the questionnaire by interviewer. This information could provide insights for training and/or question rewrite.

The authoring system allows a non-computer programmer e.g. a survey questionnaire designer, to create the questionnaire while simultaneously and automatically generating the questionnaire software component. It has been our experience that this is the most difficult component to develop. Although there are a number of such systems that are available, none of these systems has met all of our requirements for the type of complex survey we conduct e.g. NHIS. The authoring system should be extremely user-friendly and be able to handle a large number of question types.

Data Transmission

In the case of CATI, the data is automatically transmitted to a central point for either uploading to larger computer or further processing e.g. analysis.

In the case of CAPI, the data collection is dispersed generally over a wide geographic area. The two primary methods for data transmission have been mailed floppy disk or telecommunications. For data that is needed in one day or later, floppy disk has been adequate. Telecommunications, however, adds a new dimension - Two way communications. Not only can data be transmitted to a central point, but instructions for the interviewers, for example, could be transmitted from the central point to the field. The major problem with the telecommunications method has been consistent quality of the communication lines. Cost can also be a barrier.

Interviewer Training

The level and amount of training needed depends, to large extent, on the level of user-friendliness of the software. Our experience has shown that the type of training is different for either a CATI or CAPI conducted survey than for the pencil and paper conducted survey. In the paper and pencil conducted survey, training is focussed on almost entirely on the content of the questionnaire, management of the questionnaire, and the proper question sequencing. It would not be unusual to have an accompanying instruction manual 3-4 inches thick that would have to be learned by each interviewer. Whereas, in the CATI or CAPI conducted survey, training included both questionnaire content and the care and use of the computer. The major focus, being the computer not the content because the computer software can handle most of the problems the interviewer needs to worried about in the pencil and paper conducted survey, such as; probes, question sequencing, completeness.

There is one major difference between CATI and CAPI that impacts on the training: the level of interviewer anxiety. CATI is conducted at a central location where supervision and help are readily available. CAPI, on the other hand, is conducted in the field where no supervision or help is readily available. Therefore, CAPI training must try to provide the interviewers with sufficient confidence in the software and hardware to cope with this lack of help. One method that has proven effective is to emphasize hands-on practice. Interviewers are encouraged to take home their computer and practice interviews with anyone they can get prior to going into the field. In addition, interviewers are given their computer prior to the training so they can have some familiarity with them. CAPI interviewers must be able to cope with

problem occurrences. Consequently, training must concentrate on such situations.

Future Technology

Impending technological advances can have a profound impact on these automation methods; particularly CAPI. Changes in hardware such as; an "etch-a-sketch" microcomputer and an inexpensive long-life, light-weight battery would open new possibilities for the CAPI conducted survey. Use of a light-weight computer, under 5 lbs, no key board, with light pen hand-written entry would allow door step surveys as well as reduce training efforts. The "etch-a-sketch" computer has been introduced by one vendor and several other are about to announce. The long-life light-weight inexpensive battery, although not currently announced or available, when available will produce much faster and larger light-weight computers. Thus allowing larger and more complex surveys to be automated.

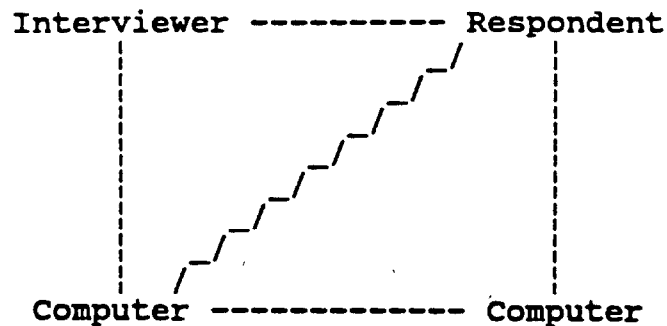
The development of an generalized authoring system software would open up the use of CATI and CAPI to the quick-turn-around type survey. Survey questionnaires could be designed and implemented quickly and easily. Staff productivity would also increase significantly because computer programming efforts to automate each survey questionnaire would be reduced to a minimum. The survey designer, in effect, would be programming the survey while designing the questionnaire.

COMPUTER ASSISTED SELF INTERVIEWING

Ralph Gillmann
Energy Information Administration

The phrase "computer assisted self interviewing" (CASI) covers all survey methods in which respondents access computers. These methods include "computerized self administered questionnaires" (CSAQ) and "prepared data entry" (PDE) where the respondent fills out a computerized version of the survey instrument. Also included are methods where the respondent uses a telephone to access a computer: "touch tone data entry" (TDE) and "voice recognition data entry" (VRE).

Let's step back for a moment and look at different ways that computers can be used in interviews:



The top line represents direct interaction between an interviewer and a respondent. The left line represents the interviewer accessing a computer such as in CATI and CAPI which were previously discussed. CASI methods are illustrated by the lower right triangle. The diagonal represents respondents accessing an agency computer as in TDE and VRE. The right line represents respondents accessing their own computers as in PDE. With the personal computer (PC) becoming ubiquitous, at least in establishments, respondents usually have access to a computer.

The bottom represents computer to computer interaction for data transmission. The missing diagonal would represent the activities of hackers and spies.

Next, let's compare manual and computer assisted methods:

<u>Survey Part</u>	<u>Manual</u>	<u>Computer Assisted</u>
Survey Instrument:	Paper	Electronic
Interviewer:	Person	Computer
Data Edits:	Manual	Automatic
Data Transmission:	Mail	Network
Data Entry:	Key In	Load File

Some methods are part manual and part computer assisted. For instance, CATI and CAPI combine a personal interview with an electronic survey instrument. One survey which uses all of the computer assisted methods is the Petroleum Electronic Data Reporting Option (PEDRO) in use at the Energy Information Administration. In general, the manual methods are slower and more prone to processing errors. Labor and postage costs are also rising faster than the operational expenses of computer assisted methods.

For transmission of the data to the collecting agency, paper copies can be sent via facsimile machines (fax). This method is faster than the mail but doesn't eliminate the need to key in the data. If the data are in electronic form, a diskette with the data can be mailed in. This is useful if security and authenticity are a particular concern. Transmission time may be saved by sending the data over the telephone network or using "electronic mail" over a computer network. (Note that it's becoming harder to tell telephone and computer networks apart.)

The use of an electronic mail service is feasible now and likely to be more important in the future. This method allows a third party to handle the support for telephone lines, security, and temporary storage. Respondents only need to have a terminal which operates over ordinary telephone lines if the survey instrument resides with the electronic mail service in the form of an electronic questionnaire. Security can be provided by passwords and data encryption. The survey agency can retrieve the data at its convenience.

Finally, CASI offers several quality improvements:

Increased timeliness of the data (especially important in monthly and weekly surveys)

Fewer follow-up calls to respondents (because many, if not all, data edits can be done immediately)

Reduced respondent burden (fewer persons are needed to fill out an electronic form)

Lower costs (at least in cases where labor and postage make up a large part of the costs)

COMPUTER ASSISTED SELF INTERVIEWING:
RIGS AND PEDRO, TWO EXAMPLES

Ann M. Ducca
Energy Information Administration

I am going to talk about two systems that the Energy Information Administration has for reporting data using personal computers (PC's). One system is a mail submission of a PC diskette, and the other uses telecommunications between the respondent's PC and our mainframe computer.

The first example is the Reserves Information Gathering System, known as RIGS. It is a system for reporting data on domestic oil and natural gas reserves on PC diskettes. The data are collected by the EIA in its annual survey of oil and natural gas well operators. Reporting to this survey is mandatory.

Briefly, this survey is a stratified sample survey with the stratification being done according to the amount of production of oil and natural gas. Respondents in the first strata, representing the largest amounts of production and having the most data to report, are eligible to report using RIGS. They will also continue to have the option of reporting on paper forms. The EIA cannot require an electronic form of submission. RIGS first became operational for the reporting of 1988 data. We anticipate that 25-30 percent of the 1989 reserves information will be reported using the RIGS system.

The EIA sends PC diskettes containing the RIGS processing software by mail to respondents. A user's guide is also provided. The respondents install RIGS onto their PC's and use it to enter data.

The basic hardware requirement is an IBM compatible PC with at least 360K of random access memory, and two floppy disk drives or one floppy and one hard disk drive. A printer should also be attached to the system so that a hard copy can be printed. Version 2.0 or higher of MS DOS is also required. The IBM PC compatible computer was chosen because of its wide availability.

The software for RIGS was originally written in dBASE III, a PC database management system. dBASE III programs can only be executed using the dBASE III software, that is, stand-alone programs cannot be created. Since the EIA did not want to purchase and provide the dBASE III software for every respondent, Clipper, a linkage compiler, was used to compile dBASE III into object code to make it a portable system. The licensing agreement with Clipper permits run-time programs created by it to be operated outside the agency. Thus, the respondents are provided with an executable load module, not programs. Licensing agreements must be carefully

reviewed before planning to use software products outside an agency.

An advantage of a load module is that respondents cannot directly or inadvertently change the programs. Also, there is no cost to the respondents since the RIGS software was developed by the government.

Using the RIGS software, the respondents enter data directly on their PC. The data entry screens for RIGS are formatted like the data collection form. There may be some benefits to exploring other formats which take advantage of options available to automated collection, such as question sequencing.

There is also the option of sending an ASCII file to the RIGS system so that data already available in an automated form at the respondent site can be submitted without re-keying. The RIGS User's Guide gives the instructions and record layout requirements for downloading ASCII files.

Respondents are required to submit to us by mail a diskette containing a copy of the cover page and the data. They must also return a paper copy of the cover page with the signature of the certifying official.

Because the survey is an annual one, it was decided that telecommunications with the EIA mainframe computer was not needed, and that the mail submission would be sufficient. Since the data in the RIGS system are proprietary, it was also decided that respondents would not be provided with their previous year's data because of the risk of sending confidential data to the wrong respondent.

Preliminary edits such as range checks are performed as the data are entered into the RIGS system. If the system detects an incorrect entry, the bell sounds and a message appears across the top of the data entry screen. The message will prompt the user for a response. Help screens are available to assist the user, and help is also available by telephone on a toll-free number. For data that have been downloaded into RIGS, an edit report is produced afterwards. A respondent may then use the RIGS edit function to correct the errors.

Final edits, such as comparisons with previous year's reports, are made after the data are returned to the EIA. These edits are performed on our mainframe system. When questionable data are identified, a quality control analyst contacts the respondent by telephone and changes are made by the EIA.

Respondents also have the option to make notes in a footnote. These notes may be helpful in explaining data that appear to be questionable.

The second example is the Petroleum Electronic Data Reporting Option (PEDRO). It gathers monthly data for petroleum supplies from petroleum companies. The respondents eligible to use PEDRO participate in 7 monthly surveys. They include refineries, storage facilities, pipelines, importers, and extraction facilities. Reporting to these surveys is also mandatory. But again, the EIA cannot require an electronic form of submission.

The participation in PEDRO varies among the 7 surveys. The market share represented by reports to PEDRO ranges from 25 to 90 percent of the total volume for a survey.

The main difference between the PEDRO and RIGS systems is that PEDRO uses telecommunications to transmit data directly to the EIA mainframe computer. PEDRO users need an IBM compatible PC with a hard disk and a floppy drive, and a modem. As with the RIGS system, respondents are provided with an executable load module at no cost. PEDRO also requires the Arbiter communications software which is licensed only for use with the EIA. Arbiter was selected because it satisfied our security needs. The EIA supplies the respondents with Arbiter.

The basic methods of entering data to PEDRO are the same as those with RIGS -- keying on the PC or sending an ASCII file to the PEDRO system. However, data submission in PEDRO is done by telecommunications directly to our mainframe, rather than by mailing diskettes. Since these are proprietary data, PEDRO submissions are encrypted. The transmissions are time-stamped to replicate a postmark. The respondents must use passwords to transmit data, and the password, rather than a written signature, serves as the certification of the validity of the data.

All edits in the PEDRO system appear on the respondent's PC. Since there is a direct link to our mainframe, all data needed for editing comparisons, for example prior month's data, are available on-line. Preliminary edits are performed before respondents transmit any data. Final edits are performed after the link to the EIA mainframe and transmitted back to the user.

The EIA is very pleased with the RIGS and PEDRO reporting systems. We believe that we are getting data faster and more accurately from these systems, and are encouraged by the increase in interest in using them.

Computer Assisted Self Interviewing

	<u>RIGS</u>	<u>PEDRO</u>		<u>RIGS</u>	<u>PEDRO</u>
Hardware	IBM Compatible PC Hard disk & 1 floppy or 2 floppy drives	Same Hard disk & 1 floppy Modem	Survey Respondents	Annual Oil & Gas Reserves	Monthly Petroleum Supplies
Software	Executable load module MS-DOS 2.0 or higher	Same MS-DOS 3.0 or higher Arbiter (R) communications	Participation	Operators of Oil & Natural Gas Wells 25 - 30 percent	Petroleum Companies 25 - 90 percent
Data Entry	Keying on a PC or Send an ASCII file to a PC	Same Same	Edits	Preliminary edits in RIGS software and Final edits at Agency	All edits in PEDRO software
Data Transmission	PC diskette mailed to agency	Respondent PC linked to agency mainframe	Comments	Option for remarks in footnotes	Same
Certification	Signed cover page	Passwords			

DATA COLLECTION

Cathy Mazur
National Agricultural Statistics Service

In this session, I will first mention several factors to consider when deciding on a mode of data collection. Then I will spend a few minutes comparing the modes of data collection that have been discussed.

The primary factors in choosing a method of data collection for a given survey are (as previously mentioned) the available time frame, the desired quality, and the cost of resources. It is unusual to have all three of these in abundance. Therefore, tradeoffs must be considered.

Several other factors to consider which relate to survey design and operation are whether the survey is mandatory or voluntary, whether a onetime or ongoing survey is to be implemented, whether households or businesses are sampled, whether the data will be collected in a centralized or decentralized manner, whether networking of computers will be done, the sample size, and the complexity of the questionnaire.

The remaining factors to consider in automated data collection refer to the characteristics of the technology. First is the speed of the hardware and data transmission over the phone lines. Next is the size of the computer's memory, and the system's weight (as in CAPI). Portability is a concern to data collection when different hardware and/or software is to be used (as in Prepared Data Entry (PDE)). The type of display is important in some modes (as in CAPI). The mode of data entry can be through the keyboard, a pushbutton phone, or using one's voice. Data verification depends on the desire for quality, the complexity of the data, and other factors. The database generation is also an important step (as was discussed by Martin Baum). It refers to integrating the data with other survey processes (label generating, data summaries). Hardware is selected based on cost, the amount of time available, the data quality desired, and the background of the staff that will operate the machines. Lastly, training is important in any survey, the amount of which depends on the technology chosen.

The priorities that are given to these factors and the relationships between them, help to decide which technology to use. All combine data collection with data entry, and most add editing at the time of data collection. This reduces the time component and increases the quality component. Also, mixed modes of data collection are possible in a survey.

First, (as a means of comparison), a mail or manual survey would require a fairly long time to send out personal enumerators or to send and receive questionnaires through the mail. The amount of editing is very limited as data entry and editing is done after all the data is collected and the interview is completed. The cost is fairly high if personal interviews are done, and nonresponse may also be high if questionnaires are mailed out.

CATI is used because it collects data quickly and accurately. The cost component (which is fairly high), comes from the hardware, software, training, and support factors (such as phone charges). One cost component which is eliminated is the travel expense. One suggestion is that CATI improves the cost benefit. The respondent, however, must have a phone. Other benefits are that it is useful in complex survey environments, can provide information on call scheduling successes/failures, and can be used for non-response follow up.

CAPI also has fairly high costs, but it provides accurate data with a tendency for higher response rates (which may be a problem in CATI), and saves on the separate keyentry time. the largest cost component is due to travel (with some in hardware and software support costs). The weight, battery life, and screen visibility are important issues to CAPI.

As to computer-assisted interviewing, 3 data collection modes are discussed -- Prepared Data Entry (PDE), Touchtone Data Entry (TDE) and Voice Recognition Entry (VRE). PDE provides faster and more accurate data, for an average cost. Costs are incurred in software development and support areas. This mode requires the availability of a PC (usually by establishments), and two issues are data security and data integration (as different PC's are used).

TDE allows respondents to call and answer questions posed by a computer using the keypad of their touchtone telephone. VRE also allows respondents to call and answer questions posed by a computer, but the respondent answers by speaking directly into the telephone, and a computer system translates the incoming sounds into text. TDE and VRE offer low cost alternatives in a short data collection time, but editing is more limited. In both, surveys tend to be shorter and simpler, non-response prompts are used, and respondent acceptance is a concern. TDE requires access to a touchtone phone and service, where VRE can use any phone. The Bureau of Labor Statistics collects data monthly for the Current Employment Statistics Program using mail, CATI, TDE, and VRE. The VRE system recognizes any American English-speaking person with continuous speech of the numbers 0-9, yes, and no.

These are not simple issues, and there are no clear cut answers. The definitions and importance of the factors must be

agreed upon. This comparison only represents the current state of technology, much will change with future development.

Lastly, I hope this session has made you more aware of the possibilities, the issues, and what to consider when choosing a data collection method.

DISCUSSION

Robert N. Tinari
U. S. Bureau of the Census

I want to begin my remarks today by noting that this paper is a very thorough treatment of the issues surrounding automated survey collection methodologies.

I am impressed with the organization of the paper and the thoroughness of discussion of the many considerations that go into selecting, designing, and implementing these types of data collection systems. The subcommittee is to be commended for the excellent job they have done in bringing together in one document a tremendous amount of information that I think will be extremely useful to those considering alternative data collection methodologies.

Based on my experience as a program manager responsible for the initial development and implementation of CATI on the National Crime Survey, there are several issues raised in the paper that I believe need more emphasis.

The first issue I want to discuss has to do with organization and its affect on CATI/CAPI development and implementation.

In its conclusion, the committee notes that increased reliance on software development has important implications for hiring and training skilled survey designers. It also states that previously distinct boundaries between occupational groups will continuously blur and disappear and survey design will likely be increasingly accomplished through teams of skilled workers from different occupations.

Based upon my experience, I believe that this is an accurate assessment. Obtaining the maximum benefit from these data collection methodologies requires that a fully integrated system be developed and this, in turn, requires the concerted effort and collaboration of programmers, survey design experts, statisticians, field staff, program managers, and survey sponsors.

However, the level of cooperation and communication necessary to successfully design and implement CATI/CAPI may be very difficult to achieve in a large, hierarchical organization. Staffs tend to be highly specialized and not experienced in projects requiring a multi-disciplined approach.

From my own experience working on one of the first CATI applications at the Census Bureau, we had a very difficult time organizing the right team with the right experience necessary to get the project underway and in keeping the lines of communication

open among the various divisions involved to implement it successfully.

We learned a lot from that process and have come a long way. A recent example is a cooperative effort between the Economic Area and the Demographic Area in successfully developing and implementing a CATI system for the Survey of Manufacturing Technology. The Industry Division was responsible for conducting the survey and wanted to use CATI for nonresponse followup of manufacturing plants. The division lacked the experience to develop the questionnaire on CATI. Demographic Surveys Division offered to help with the authoring, Industry assisted with testing and Field Division worked on interviewer training and data collection. The survey was carried out on time, within budget, and with high quality. This is a good example of what can be accomplished by individuals working together from the various divisions and sharing their expertise to get the job done.

Poor organization and control can have a very serious impact on the cost and time of development and the quality of the final product. I believe that what is needed to successfully design and implement automated data collection methodologies is:

- o commitment and full support from upper-level management.
- o a full-time, dedicated staff - no part-time work along with other projects.
- o open lines of communication with clear assignment of responsibility/accountability.
- o designate a project coordinator/facilitator
- o breaking down of traditional barriers between survey statisticians, mathematicians, survey designers, programmers, and field staff in order to work effectively.
- o ongoing commitment and organizational change to adapt to needs of the new data collection methodology. Especially important if you are using mixed mode such as personal visit (paper) and centralized telephone (CATI).
- o reduced layers of bureaucracy.
- o empowerment of the team to get the job done.

We must think of new ways of organizing ourselves to be more flexible and effective in designing and implementing new technologies. In addition, there must be more sharing of

information among the various statistical agencies on approaches and experiences in the area of organization.

The second issue has to do with interviewer acceptance of new technologies like CATI and CAPI. The paper points out the importance of involving the user in the design process. I do not think this point can be over-emphasized.

In the rush to develop survey instruments on tight time schedules or in deciding which portable machines to use for CAPI applications, we the developers and/or program managers, take it upon ourselves to decide what is best for the interviewers and may not actively involve them in the decision or development process. This can be a big mistake.

If the interviewers are not comfortable with the interface, if it is slow, clumsy or awkward to use, "not natural" feeling, not helpful, etc., the survey is in serious trouble. If the interviewers have no say in the design and for any reason should decide that the system is not helping them to get the job done better, then you face an uphill struggle to gain their acceptance, and in some instances, the system may never be fully accepted. Interviewers may work to defeat the system, morale may suffer, respondent cooperation may suffer, turnover rates will increase, quality will suffer, and costs will escalate.

In addition, if you are contemplating switching from a personal visit environment to CATI, you must consider the effect on the interviewer staff out in the field. Field interviewers will be concerned about losing their jobs and quality may suffer during the transition to CATI. How the Field interviewers will be treated and possible impact on data quality during the transition period should most definitely be taken into account. For example, in planning the transition of cases from personal visit to CATI for the National Crime Survey, we used attrition among interviewing staff and hard to enumerate areas for conversion to CATI. By using this approach, CATI was viewed as positive tool by Field staff. This plan helped to gain acceptance of CATI.

The third and final area I want to discuss has to do with the need for adequate testing and evaluation of these new methodologies.

Before implementing any survey operation, it is good practice to allow enough time for adequate testing and evaluation of the instrument and the data collection and processing system. This is especially crucial for automated data collection systems. Complex questionnaires (those with complex branching or edits) need to be thoroughly tested and evaluated before they are introduced on a production basis.

While the automated data collection systems provide us with the ability to field much more complex questionnaires than we could using conventional paper forms, they also pose additional challenges related to testing. Aside from the obvious problems that may surface during interviewing, if the instrument is not adequately tested, there may be logic errors hidden in the instrument that go undetected or aren't found until after the data collection phase is complete.

In addition, when changes are introduced to the questionnaire, (even minor ones), thorough testing should be conducted again to insure that other questions or skip patterns have not been affected.

In the paper, the committee discusses the possible application of expert systems in questionnaire development. I would suggest that perhaps some application could be found for these systems to testing and evaluating as well. There is definitely a need for more systematic and thorough methods for checking out the questionnaire. In addition, attention must be paid to testing the case management, call scheduling, training, data transmission, and processing systems before the survey is fielded.

This is not something that only needs to be done before a survey is fielded. It should be an ongoing effort to evaluate how well the system is functioning. It should allow for feedback for continuous improvement/refinement such as monitoring, observation, debriefing interviewers/respondents.

I want to thank the organizers for giving me the opportunity to share my views on this important topic. I think the committee has made an important contribution by bringing together in one document many of the issues facing project managers in deciding whether or not to adopt these technologies. I hope that the document will be treated as a dynamic one that will be expanded as we gain more experience with the various aspects of these data collection methodologies.

DISCUSSION

David Morganstein
Westat, Inc.

I thank Terry Ireland for organizing this intriguing session and I would like to express my appreciation to the speakers for the work they have done in their examination of new methods for assisting in the process of conducting government surveys. It is a pleasure to be given this opportunity to participate in the session as a discussant.

The job description for a discussant might be:

- To agree with the speakers comments,
- To point out errors or omissions,
- To suggest areas of new research, or
- To do something completely different that they'd like to do!

I think I will try a little of all four of these objectives.

There is a great need for new approaches to gaining cooperation as the respondent population is increasingly bombarded with requests for survey participation. The initial 1990 Census experience indicates the level of difficulty surveyors can anticipate.

According to our speakers, their "Primary job ... is to develop ... computer related techniques which allow the respondent to answer the survey completely and accurately". The emphasis on the respondent's cooperation is very appropriate. There is a potential trap of having the software developed by software experts who have little knowledge of or interest in the respondent/interviewer who must use the system. At a minimum, a part of the system designer team should be practitioners of long standing who understand the process. There may be good reason to have this leader of the team be such a practitioner.

I was concerned by the following statement found in the paper, "Interviewers must believe that Computer Assistance will improve their effectiveness. They need to be convinced that the computer is simply a tool to expedite and simplify their work." This sounds a bit like psychological behavior modification. Such verbal persuasion should be unnecessary. In fact, the users WILL believe and be convinced IF the system actually DOES this! You can be sure that no amount of argumentation will insure the interviewer's support if the system is awkward, difficult to use and makes their work harder.

The focus of the paper was primarily on the technology. It said little about comparison studies which measure the accuracy/reliability of CASI responses as compared to more traditional methods. For example, an 1984 paper by Waterton & Duffy in the International Statistical Review indicated self-reported alcohol consumption rates that were significantly higher when obtained via CASI than previously measured by interviewer. Perhaps there have not been enough such studies, however, there is a need for them.

The paper pointed out the importance of a good authoring system to CAPI but didn't say the same for CATI. I believe it is true in that environment as well.

Quality Measures (Human Interface discussion) are very important and are needed if we are to evaluate the efficacy of these new approaches. The authors also mentioned an evaluation by 'user' (interviewer), something I agree is important as it speaks to the committees 'primary job' mentioned earlier.

I found the Appendix 3 examples a useful reference for contacts. The authors would perform a valuable service if they would include names and phone numbers for all contacts.

These approaches conform to the modern concept of quality. Reduced variability is designed into the system. They reduce the potential for 'creative interviewing' in which undesired variation is introduced by the interviewer during the interview process.

While I have not worked with CASI, it would appear that it could suffer from a potential loss of control by the survey operator. It could be subject to 'creative respondents' who are intrigued by technology or who seek to befuddle the survey operators. Care must be taken to insure that this does not occur.

The survey instrument's logic/design still depends upon the human mind. Techniques for encoding it into a CATI/CAPI/CASI system need to be better understood. An unrealized advantage of these methods is that they force the designer to better understand the instrument/flow earlier in the process. The designer can't rely upon last minute training/role plays with the interviewers to clarify muddy logic or instrument flow.

I would like to close my comments on the value of these high tech methods for assisting in survey operations with the following short essay on the beauty of the abacus written by Robert Fulghum.

Essay taken from All I Needed to Know I Learned in Kindergarten, Robert Fulghum.

Session 7

QUALITY IN BUSINESS SURVEYS

IMPROVING ESTABLISHMENT SURVEYS AT THE BUREAU OF LABOR STATISTICS

Brian MacDonald
Alan R. Tupek
U. S. Bureau of Labor Statistics

Introduction

The report on "Quality in Establishment Surveys" (see Statistical Policy Working Paper 15, 1988) concluded that there were few commonly accepted approaches to the design, collection, estimation and analysis of establishment surveys. In contrast to household surveys, there was little standardization of methodological approaches across establishment surveys. The report classified potential sources of errors in establishment surveys and examined the range of practices which are used to improve and measure quality.

Each Federal agency which collects statistical data from establishments develops their own frame of business establishments. These frames are of varying quality, which greatly affects the methodology for surveys and contributes to the divergence of methodology across establishment surveys.

This paper first provides a summary of the design considerations for establishment surveys as discussed in Statistical Policy Working Paper 15. This paper then describes the efforts at the Bureau of Labor Statistics (BLS) for improving their business establishment list, the effect of these improvements on BLS surveys, and the potential impact on other statistical agencies.

Design Considerations for Establishment Surveys

Establishment populations differ from household populations in several ways (see Statistical Policy Working Paper 15). These dissimilarities result in frame development, sample design, and estimation approaches which are in some areas markedly different from approaches for household surveys. Among the major distinctions between establishment and household populations and frames are:

1. Establishments come from skewed populations wherein units do not contribute equally (or nearly equally) to characteristic totals, as is the case for households; and
2. Accuracy of frame information about individual population units is crucial to sample design and estimation for establishment surveys, while for household surveys the

accuracy of frame characteristics concerning individual units is not as critical to the sample design.

Establishment surveys are characterized by the skewed nature of the establishment population (see, for example, Table 1). A few large firms commonly dominate the estimates for most of the characteristics of interest. This is especially true for characteristics tabulated within an industry. Small firms may be numerous, but often have little impact on survey estimates of level although they may be more critical to estimates of change over time or for measuring characteristics related to new businesses. This distribution has a major impact on both the frame development and maintenance and on the sample designs used for establishment surveys.

Table 1

Distribution of Establishments on the Bureau of Labor
Statistics List Frame by Number of Employees
(First Quarter, 1989)

SIZE CLASS (No. of employees)	% OF ALL UNITS	% OF ALL EMPLOYEES
ALL	100.0	100.0
0 - 4	58.3	6.2
5 - 9	17.9	7.6
10 - 19	11.1	9.7
20 - 49	7.6	14.8
50 - 99	2.8	12.4
100 - 249	1.6	15.9
250 - 499	0.5	10.0
500 - 999	0.2	8.0
1000+	0.1	15.4

SOURCE: U.S. Bureau of Labor Statistics

List frames are widely used in establishment surveys conducted by the Federal government. The use of list frames for establishment surveys arose from the availability of administrative records on businesses compiled mainly for tax purposes. However, because these administrative record files are not normally developed for statistical purposes, they often need refinement before being used as sampling frames for surveys of businesses. Extensive resources are spent on maintaining the list frames since a significant source of nonsampling error may be due to inadequacies in the frame.

Establishment list frames typically are characterized by detailed establishment identification information, periodic updating of this information, and multiple sources for the information. The data on the frame are required for sample design, sample selection, identification of sample units, and estimation. The primary source of administrative records for a frame may have shortcomings which require the identification information to be supplemented using other sources of information. This may include using identification information from the surveys themselves. Supplemental files, including the use of area frames, may also be required to overcome coverage problems in the primary source. Duplication of sampling units may also be a problem associated with the use of list frames. Refinement of the frame includes efforts to unduplicate units prior to sampling.

The individual establishment information on the frame is critical to the effectiveness of the sample design and estimation for the survey. Maintaining a frame over time is complicated by the dynamic nature of the establishment community. Changes in ownership, mergers, buyouts, and internal reorganizations make frame maintenance a real challenge. Matching and maintaining unit integrity over time provides the opportunity for consistent unit identification in the numerous periodic surveys conducted by the Federal Government.

New establishments must be added to the frame. However, it is often difficult to differentiate, using administrative records, new establishments from formerly existing establishments that have changed their name or corporate identity. It is also difficult to link businesses over time when there have been ownership or other changes. Each survey may have different requirements as to the handling of new establishments and changes in existing establishments. The timeliness of adding new establishments to the frame and reflecting them in the sample is also a problem. The lag time between formation of new establishments and selecting them into the sample may be anywhere from several months to several years. While new establishments may have little impact on estimates of level, in some instances they may dominate estimates of change.

The Business Establishment List Improvement Project

In May 1987, the Economic Policy Council issued a report that noted five areas in national economic statistics where improvements were needed. One of these areas dealt with the business lists used by the three major Federal statistical agencies to conduct their surveys. One of their recommendations was that the Bureau of Labor Statistics and the National Agricultural Statistical Service of the Department of Agriculture be designated as the central Federal government agencies for the collection of nonagricultural and agricultural, respectively, business identification information.

In addition, the Economic Policy Council recommended that efforts be initiated to revise the statutes that prohibit the sharing of survey data collected by the Census Bureau with other specified Federal statistical agencies. The main purpose of the Economic Policy Council recommendations was to have a single, high-quality source of business data available to selected Federal statistical agencies in order to increase the quality and comparability of national economic statistics.

Shortly thereafter, the Office of Management and Budget (OMB) requested that the BLS develop a proposal to assume this role. The issue of devoting resources to developing a central frame is not unique to the fragmented U.S. statistical system. Statistics Canada is in the process of developing a central frame for its business establishment surveys (see Colledge and Lussier 1987).

For the BLS universe file to sufficiently serve as the primary frame for statistical survey sampling by Federal statistical agencies, the BLS recognized that modifications to its existing file were necessary. The most critical need was to improve the information available about employers engaged in multiple operations within a State. The Business Establishment List (BEL) Improvement Project was initiated to do this. Its primary purpose is to create an establishment (i.e. worksite) based register of units with full identification information on United States' businesses. At present, data for multi worksite employers in the BLS register are available mostly at a higher level of aggregation.

The data for the current BLS universe file come primarily from administrative records collected by State Employment Security Agencies (SESAs) as part of the administration of the Federal/State Unemployment Insurance (UI) System.

All employers covered by unemployment insurance are required to file quarterly UI Contributions Reports with the SESAs for each of their UI accounts. On these forms, employers report the number of full and part-time workers, employed during the pay period including the 12th of each month in the quarter and the total payroll for the quarter. This reporting is mandatory for single location employers as well as those engaged in multiple operations in the State.

Data collection and classification procedures for multi-unit employers differ from those for single units. For multi-unit employers, the statistical branch of the SESA is responsible for the direct collection and review of monthly employment and quarterly wages at the reporting unit (county by industry) level of detail. A multi-unit employer is defined as an employer who has more than one industrial activity (four-digit SIC) and/or county location covered by the same UI account and meets the following criteria. To qualify as a multi-unit employer, the employer must have 50 or more employees in the sum of their secondary industries

or counties. The primary industry or county is defined as the industry or county that has the greatest number of employees.

Under the BEL Improvement Project (see Searson and Pinkos 1990), this threshold is being lowered from 50 employees to 10 employees with the States being responsible for collecting employment, wage and identifying data at the worksite level. Thus, more detailed business identification information will be available for small multi-establishment employers.

Multi-unit employers that do not meet the above criteria are treated as if they were single-unit employers for data collection and recordkeeping purposes. These small multi-unit employers who are engaged in multiple industrial activities within one county are assigned industry codes based on their primary activity (that is, the activity providing the most shipments or sales). Conversely, those in one industry with several locations are given a county code based on the location employing a majority of all the employees.

Large multi-unit employers are treated differently than single units as they are requested to file a quarterly statistical supplement form in addition to the Contributions Report. On the SESAs' current forms, large multi-unit employers report monthly employment, quarterly wages, industry and location information for each reporting unit. These supplements are used to maintain separate identification and characteristic records on the individual reporting units to ensure correct geographical and industrial totals are maintained.

As part of the BEL Improvement Project, the BLS is replacing the 53 individually-designed State forms with a standardized statistical supplement form. The name of the form is being changed to the Multiple Worksite Report. Each quarter, the employer will be requested to verify the identifying information (trade name, description of the establishment, and physical location address) for each establishment (worksite) that will be computer printed on the new Multiple Worksite Report. In addition, the employer will be requested to provide the monthly employment and total wages for each worksite for that quarter. By using a standardized form, the reporting burden on many large employers, especially those engaged in multiple economic activities at various locations across numerous States, should be reduced. States will accept listings and floppy diskettes of this information in lieu of the form. In addition, the BLS is investigating the central collection of multiple worksite employers data from major multi-establishment employers. The Multiple Worksite Report form will be used in all States to collect data by establishment (worksite) beginning with data for the first quarter of 1991. Some twenty-one States, however, are switching to a State version of the new form with data collected for the first quarter of 1990.

As a result of these efforts at worksite reporting, we expect the number of units on the frame to increase from approximately six million to slightly more than seven million. Because the UI system still serves as the basis for the worksite based frame, both the scope as well as the data on employment and wages on the new frame will be identical to that on the old frame, only the level of disaggregation will be different.

Implications of BEL on BLS Surveys

Several features of the BEL Improvement Project will affect the design of BLS sample surveys (see Plewes 1989). These include:

- reporting unit number for each worksite of multi-unit companies;
- better identification information, including multiple unit multiple addresses, worksite descriptions, and telephone numbers;
- better linking of data over time through the use of reporting unit number for worksites within multi-unit UI account numbers. Also, UI accounts will be linked through the use of predecessor and successor codes for ownership changes such as buyouts, mergers, etc;
- more data items for each unit, such as initial date of tax liability, date of establishing a new worksite, and comment codes for explaining unusual employment changes;
- quarterly data, historical files, and response history files to track the surveys for which a worksite has been selected and whether they have responded;
- linking of units within enterprises or corporations, across UI accounts; and
- improved standard industrial classification (SIC) refiling process, in order to identify new multi-worksite reporters in addition to updating SIC codes on a 3-year cycle.

The effect of these BEL improvements on four areas of survey design will be examined. These include sample frame development, sample design, data collection, and estimation. Implications for the short-term, during the period in which the survey program will transition into the improved system, as well as the long-term will be discussed. The transitional period implications are usually related to problems in maintaining consistency of survey estimates while BEL improvements are implemented. The long-term implications

are usually related to improvements that can be made to survey designs by reexamining survey design objectives.

Over the years, each BLS survey has developed activities for creating their sampling frame from the old Universe Maintenance System, which BLS will change. These unique activities for each survey focus on specific survey requirements as well as limitations of the list. For example, BLS surveys which attempt to maximize sample overlap over time must match frame units from one time period to another. The BEL improvements will affect the matching operation, due to the shift to worksite reporting. During the transition period, the surveys may need to reexamine the need to maximize sample overlap. If they maintain this objective, then less sample overlap is likely, and much of the operation will need to be done manually. However, in the long-term the use of reporting unit numbers, and predecessor and successor codes should greatly facilitate the automated matching operation. Other BLS surveys use supplemental frames to survey populations not entirely covered by the BEL. These populations may include railroads; federal, state and local government; religious organizations; and seasonal industries. BEL improvements will allow many surveys to reexamine the need for supplemental frames, especially for state and local governments, and seasonal industries.

Several other long-term benefits for sample frame development are possible through BEL improvements. The availability of quarterly data can be used by some surveys for creating their sample frame. The identification of new businesses on the BEL can be used as a stratification variable for surveys.

Although BLS does not now do so, the new list will enable survey operators to conduct surveys of enterprises or companies. This will bring about reconsideration of the scope of the surveys. All surveys will need to modify their control file systems to handle additional data items on the BEL.

At this stage of the planning process, certain obvious changes have been identified for each survey. The following three examples illustrate the types of operational modifications which are planned.

- First, the survey which is used to develop the Producer Price Index (PPI) must use 1st quarter data for measures of size. The BEL improvements will allow PPI to use more current quarterly data, or other quarters for seasonal industries. This is expected to improve the coverage of some industries, and to increase the sample design efficiency.
- Second, an annual survey which measures occupational industries and illnesses supplements the BEL with a frame of the 500 largest companies in the United States,

including all of their subsidiaries. Currently, this supplemental frame is developed specifically for this survey. The BEL improvement plan will provide adequate organizational relationships for large companies, so that the separate operation will be terminated.

- Third, a monthly survey of employers, which measures employment and average hourly earnings, lags in measuring the affect of new businesses. A sampling strategy is being developed for this survey, which will bring in a sample of new businesses each month, once the BEL improvements are introduced.

Greater flexibility in sample designs will be possible with the introduction of BEL improvements. Separate strata for seasonal or volatile firms can be considered. Stratification by age of firm may be appropriate for some surveys. Surveys designed to produce local area estimates can use worksite locations for stratification. Surveys may want to stratify by multi-reporters versus single reporters, or by enterprise size. The survey response history can be used to avoid overlap between surveys and to spread respondent burden.

During the transition period for BEL improvements, there will be some loss in sample design efficiency. The use of current data to develop sample designs for surveys conducted during the transition period will be somewhat inappropriate. In the long-term, sample design efficiencies will be possible through the use of new design variables and more homogeneity within size classes.

Surveys with size cutoffs will need to reevaluate the survey scope or target population. Some BLS surveys cover only large establishments. For example, most of the occupational wage surveys cover only establishments with 50 or more employees. The BEL improvements will shift units between size classes. In general, the sampling unit will shift from a county-wide report to a worksite report. Maintaining a 50 or more employee size cutoff will artificially move units in or out-of-scope of the survey and decrease employment coverage. The effect on wage estimates will need to be examined, and decisions made on how to maintain consistency over time.

Surveys designed to measure change can use the linking of data over time to improve on the efficiency of the sample design through sample overlap. Samples for surveys conducted three or more years apart are now independently selected. With historical relations maintained over time, samples could be selected which improve upon estimates of change, possibly using composite estimation.

The new features of BEL will be most beneficial during the data collection phase. Because of better address information, especially physical location addresses and telephone numbers,

response rates are expected to increase for mail and telephone surveys, since one of the primary reasons for low response rates is failure to reach the correct respondent. Additionally, better address information will result in a decrease in data collection time and effort, such as reduction in telephone and mail follow-up of nonrespondents.

The breakdown of the multi-establishment companies that presently report on a consolidated basis (e.g., county-wide) into establishment or worksite level reporting will affect all BLS surveys. Surveys will need to make special reporting arrangements with these companies to provide data on a worksite basis. Recent cognitive research conducted by Statistics Canada shows that respondents who are in the survey on a regular basis report data in the same manner from one time period to another and usually do not take into account changes to the survey instrument or procedures. The worksite information should reduce the reporting error due to failure to identify the selected sample unit.

The impact of BEL during the estimation process for BLS surveys will vary significantly by survey type and estimation procedures used. An area of survey estimation that will be affected by BEL is benchmarking. Benchmarking is a process that accounts for changes that occur during the time lapse between the reference date of the sampling frame and the date of data collection. In other words, it accounts for births, or those units which have come into existence since the sampling frame was created. This is accomplished by multiplying the sample estimates of totals by the benchmark factor at the estimating cell level, usually SIC or size class within an SIC. For BLS surveys, the benchmark factor is calculated at the estimating cell level as the ratio of the reference period employment (benchmark employment) to the weighted employment from the sample.

Surveys that benchmark at the size class level would be most affected because of the change in the distribution of units across size classes due to worksite level reporting. For example, size class benchmarks for a survey that measures occupational employment statistics (OES) by industry may be inappropriate during the transition period. A possible solution for all surveys which benchmark by size class is to benchmark at the industry level during the transition period.

With the new business registry, population data for benchmarking employment will be available for all 12 months. This additional information may be utilized by the Current Employment Statistics (CES) Survey, which is a monthly survey of about 300,000 establishments that measures employment at National and State levels by industry, to benchmark the employment data quarterly and thereby better analyze the components of error by time period.

Central Agency Status

When the OMB issues the directive naming BLS as the central agency charged with maintaining a list for nonagricultural businesses, several actions will have to be undertaken before extracts from the BLS list can be made available to other Federal statistical agencies for use in surveys.

First, BLS will have to conduct a series of negotiations with the State Employment Security Agencies to gain their agreement to waive or modify existing State confidentiality rules and regulations that would currently not allow widespread use of the state provided UI data. We expect that most SESAs will readily welcome the sharing for statistical purposes of these data. There have recently been examples where most, if not all, State agencies authorized this type of data sharing, but on a much more limited basis. In those few States where current State law might prohibit the sharing with other Federal statistical agencies, we will propose modifications to the State Unemployment Insurance laws to allow the sharing and work with the state agencies to seek passage of the needed legislation.

Similarly, there will have to be certain actions taken both by BLS and those Federal statistical agencies authorized by OMB to have access to the BLS list before the sharing can begin. BLS will have to develop formal procedures for use of the file by other agencies. These procedures will include such obvious items as security measures for the data, assurances that the confidential data will be used for statistical purposes only, agreements on feeding back 'corrections' or updates to the file, access rules and techniques (the BLS list is maintained at the NIH computer facility) and arrangements made to cover marginal operating costs for providing the data. A possible solution to the question of providing for satisfactory computer security may be for the using agency to have conducted an application security review for its own sensitive Automated Information System in compliance with the requirements of OMB circular A-130.

Summary

A central agency charged with maintaining a list of nonagricultural businesses provides an opportunity for improving business establishment surveys conducted by the Federal Government. However, the key to its success will rest with the ability of all the agencies involved to provide clear and concise requirements to the central agency, and to weigh the costs of improvements to the central list against the benefits to survey operations and data quality.

References

Colledge, M. and Lussier, R. (1987), "A Generalized Methodology for Economic Surveys" in Proceedings of the Business and Economic Section of the American Statistical Association Annual Meetings, pp. 131-149.

Plewes, T. (1989), "Improving the Business Establishment List: Survey Design Implications" in Proceedings of the Fourth International Roundtable on Business Survey Frames, Newport, Gwent, United Kingdom: Available through the U.S. Department of Labor, Bureau of Labor Statistics, in press.

Searson, M., and Pinkos, J. (1990), "The Bureau of Labor Statistics' Business Establishment List Improvement Project" in Proceedings of the Sixth Annual Research Conference, Washington, D.C.: U.S. Department of Commerce, Bureau of the Census, in press.

Statistical Policy Working Paper 15 (1988), "Quality in Establishment Surveys", U.S. Office of Management and Budget.

A REVIEW OF NONSAMPLING ERRORS IN FEDERAL
ESTABLISHMENT SURVEYS WITH SOME AGRIBUSINESS EXAMPLES

Ron Fecso
National Agricultural Statistics Service

Working Paper 15 (WP-15), "Quality in Establishment Surveys," addresses the accuracy of establishment surveys. Although WP-15 concentrates on accuracy, we need to recognize that accuracy is only a part of the total quality picture. Remember the importance of other aspects of quality and their interaction with accuracy concepts. The definition of survey quality is the totality of features and characteristics of a survey that bears upon its ability to satisfy a given need. Sometimes these ideas are referred to as "fitness for use." Discussions of quality usually address how well something is made. We must also address the true needs of the product or service as well as productivity issues such as increased output and unit cost. Continued pressure on budgets and demands for increased statistical output are quality aspects which may be occupying major portions of our time. Thus, a model for survey quality needs four elements: accuracy, timeliness, relevance and resources.

The intent of this paper is to provide a glimpse of the nonsampling error treatment from WP-15 and several examples of the treatment of nonsampling errors in agricultural surveys. I hope that I can persuade the audience to study working paper 15 in more detail after seeing this commercial.

Many sources of error are possible in establishment surveys. While there are several good ways to organize the presentation of these errors, WP-15 chose two main groupings: design and estimation, and methods and operations. The latter group contains the nonsampling errors which are highlighted here.

Nonsampling Errors

Errors which arise during the specifications for and the conduct of establishment surveys are called nonsampling errors. Commonly known examples of nonsampling errors include incomplete sampling frames, nonresponse and keypunching errors. The variety of nonsampling error sources and results from studies of these sources lead survey researchers to believe that nonsampling errors may often far exceed sampling error. There are three objectives found in the chapter on nonsampling errors in WP-15. The objectives are to outline major categories of nonsampling errors in establishment surveys, to identify some of the diverse sources of error in each category, and to provide insight into strategies to detect, measure, and control these errors. The error categories

discussed are specification, coverage, response, nonresponse, and processing errors.

WP-15 defines each of these error groups, gives examples, identifies major sources of the error, describes methods to control and measure the errors, and profiles the control and measurement techniques used in the major establishment surveys of the Federal Government (9 agencies and 55 surveys). (The presentation contained some detail about response error treatment and examples of WP-15's graphics since most of the audience had not seen WP-15. These materials are not reproduced here.)

Although several good references are available concerning nonsampling errors in surveys of individuals (for example United Nations, 1982), WP-15 is the first detailed treatment for Federal establishment surveys. The need for this separate treatment arises because establishment surveys differ from surveys of individuals by typically seeking hard data for which records are available. This characteristic both simplifies the collection and complicates the interpretation of the data. The collection is simplified when hard data on record can be used, rather than relying on the memory, opinions, or interpretations of the respondents. These differences present complications when establishing the concepts and definitions to be used in the surveys. Special care must be taken to consider carefully the establishments' recordkeeping systems, definitions, and data availability to avoid introducing specification error into the data.

Establishment surveys, which commonly use list frames, are subject to errors such as duplication, overcoverage of out-of-scope and out-of-business units, under coverage of business births, and misclassification of units. The availability of records affects the structure of the response and nonresponse errors as well as the methods to measure and control them. The treatment of processing errors differs the least from other types of surveys.

SOME HIGHLIGHTS OF WP-15

WP-15, unfortunately, makes no specific recommendations. Yet, the profile of nonsampling error practices used in 55 Federal establishment surveys by nine agencies provides considerable insight into the state of quality in these surveys. This commercial for the paper will present a few of the highlights.

- o No single measurement of specification error is used in a large majority of the surveys profiled.
- o Relatively little is done to measure specification error.
- o Few direct measures of list coverage error were reported as regularly used.

- o Outside of the calculation of edit failure rates, little response error measurement is done.
- o Although follow up procedures for large units are common, very little is done to directly measure nonresponse error.
- o Cognitive studies are rare.
- o Questionnaire pretesting was not widely used on a regular basis.
- o Relatively few nonsampling error measurements are published.
- o There is relatively little information about processing errors.

WP-15 contains considerably more detail on good practices which are currently in use as well as those practices which are lacking in use and need examination. WP-15 states in an overview that "Nevertheless, the tenor of the findings can be depicted as recommending more work to improve and document the quality of surveys... a need to focus additional attention, and resources, on the general improvement and documentation of survey practices."

A Reinterview Study from Agribusiness

An example of measuring response error in an establishment surveys is next. The results presented are from a reinterview study which measured the bias of Computer Assisted Telephone Interviewing (CATI) methods on a National Agricultural Statistics Service (NASS) survey. (Fecso and Pafford) As part of its estimating program, the NASS publishes quarterly estimates of crop acreage, intentions to plant, actual plantings, harvested acreage, stocks of grains, and livestock numbers. The source of these estimates is a multi-purpose, multi-frame survey.

Because of the detailed nature of acreage, stocks and livestock inventory items, the NASS had relied primarily on personal interviews to get the most accurate answers from the farm population. For example, on-farm grain stocks data, extremely important because of their effect on commodity trading, is a collection problem because farmers may store these grains in multiple bins on property they own and/or rent. In addition farmers often have multiple operating arrangements involving their own grains, those of landlords, and those where formal and informal partnerships exist.

Recently, NASS has expanded the use of telephoning, including CATI to collect these data. The primary reasons for change are

inadequate budget and the need to reduce the time between initial data collection and publication. We suspected difficulty in using the telephone to collect some of these quarterly survey data. Obtaining accurate responses is difficult because of the detailed nature of these data and the centralized (state) telephone interviewers often lacking farm experience and familiarity with farm terms. The reinterview study is our first attempt to measure response errors.

You can find the use of reinterview methods in the literature for measurement of simple response variance (Bailar, 1968; O'Muircheartaigh, 1986) and correlated response variance (Groves and Magilavy, 1986), for example. This response error study focused on measurement of the bias by treating the final reconciled response between the CATI and independent personal reinterview response as the "truth." To obtain truth measures, experienced supervisory field enumerators reinterviewed approximately 1,000 farm operations for the December 1986 Agricultural Survey. The following tables contain the results for the grain stocks items (corn and soybean stocks).

Table I indicates that the difference in the CATI and final reconciled responses, "the bias," was significant for all but one item (soybean stocks in Indiana). The direction of the bias indicates that the CATI data collection mode tends to underestimate stocks of corn and soybeans.

The process of reconciliation identified the reasons for differences. A summary given in Table 2 indicates that an overwhelming percent of differences (41.1%) could be related to definitional problems (bias related discrepancies), and not those of simple response variance (random fluctuation). Definitional discrepancies contributed almost half of the large bias. About two-thirds of the definitional discrepancies had a relative difference (the reconciled response minus the CATI response divided by the CATI response) more than 25% or less than -25%. In contrast, the differences due to rounding and estimating contributed less than 10% of the overall bias. Almost all of the rounding and estimating relative differences were between -25% and 25%.

TABLE I. Estimates of Bias in CATI Collected Responses 1/

State	Corn Stocks % of 1,000 Bu. CATI		Soybean Stocks % of 1,000 Bu. CATI	
Minnesota	60,711 (10.4%)	*	13,230 (14.9%)	*
Indiana	53,218 (17.9%)	*	2,998 (5.9%)	
Ohio	22,711 (12.0%)	*	5,434 (13.7%)	*

1/ The sampling design in the original CATI sample was stratified simple random sampling. The reinterview sample was a random sample of CATI respondents within strata. The bias was approximated by expanding the difference in reconciled and CATI response at the sample unit level.

* Indicates the CATI and final reconciled response were significantly different at $\alpha=.05$.

These results suggest that we can reduce the bias in the survey estimates generated from the CATI telephone sample using a revised questionnaire design, improved training, or a shift in mode of data collection back to personal interviews. Considering the constraints of time and budget, the change to additional personal interviews is unlikely. Thus, the alternative is to use reinterview techniques to monitor this bias over time to determine whether the bias has been reduced through improvement in questionnaires or training. If large discrepancies continue, the estimates for grain stocks can be adjusted for bias through a continuing reinterview program. If the bias stabilizes, even at zero, periodic reinterview studies can validate a "constant" bias adjustment used in interim periods.

An Example -- Bias Measurement

NASS conducts crop yield surveys in states which are major producers of field crops. The survey data are used to forecast expected yield and production during the growing season and to estimate these values at harvest.

Briefly, the survey design can be described as a multiple step sampling procedure. Samples are drawn from an area frame to estimate acreage for harvest, followed by subsampling of fields and small plots to make measurements related to yield per acre. Detailed information on the area frame design is available in Fecso, Tortora and Vogel. More detail on the crop yield surveys, called objective yield (OY) surveys, is in Matthews (1985), Reiser, Fecso and Taylor (1987), and Francisco, Fuller and Fecso (1987).

TABLE 2. Reasons for Differences in CATI and Reinterview Responses for Corn Stocks in Minnesota - December 1987.

REASON	NUMBER	PERCENT OF TOTAL
ESTIMATED/ROUNDING	28	31.1%
"DEFINITIONAL"	37	41.1%
operation crossing state boundaries		
misunderstanding of what to include/exclude		
failed to report reserve corn		
failed to include grain belonging to someone else		
confused with reporting that stored in town		
confusing with reporting landlord corn on farm		
didn't ask/forgot to include last year's crop		
included rented storage on other farms not his		
not all bins included		
rented bins not included		
bins on son's farm mistakingly incl.		
didn't ask about ear corn/only reported shelled corn		
confusion with reporting high moisture corn		
forgot to include grain purchased for feed/seed		
gave all whole grain/not just corn		
included cracked corn in silage		
didn't include storage used by another		
confusion with storage on acres operated versus where lives		
only included stocks of one partner/the one interviewed		
included only new crop in storage		
thought CATI enum wanted only gov. program corn stored		
original figure did not reflect Dec. 1 stocks		
OTHER	25	27.8%
<u>TOTAL</u>	<u>90</u>	<u>100.0%</u>

Several control procedures existed for the OY surveys. Supervisory enumerators visited the plots (approximately a 10 percent subsample which included the first sample visited by each enumerator). The field office survey statistician occasionally visited plots. Data are hand and computer edited. Finally, periodic validation surveys, covering a subset of crops and states in a given year, were conducted to measure the overall bias of the survey estimate in the domain studied.

These control procedures had shortcomings. For example, visits by the supervisory enumerator served mostly as a retraining system; the data was not used to improve the estimates or to estimate biases. Budget and staff reductions reduced the number of field visits by survey managers. Edits have been changing. New computer edits and some areas creating individualized recording forms have resulted in estimates which may differ from those based on the old editing procedures. Finally, the expensive and administratively burdensome validation survey received increased questioning.

The validation survey had one major goal -- to measure the differences between the objective yield crop cutting and the farmer's harvest. The validation surveys had clearly shown that the difference between the OY crop cutting and farmer's harvest is not equal to zero. These studies found differences by crop, year, and state. Since the validation surveys have answered the major question for which they were designed, we asked what purpose would they have in the future?

Our main consideration remained the assessment of the bias. Several concepts needed attention. Was the overall bias consistent over the years? Our data is a time series, especially when considered by the users; thus, knowledge of bias-included level change is important. Are the sources of bias changing? Are there large enough bias changes to deserve extra concern? Are there any needs for procedural changes to reduce specific bias sources, or do we only need to monitor the overall level of bias? Finally, if we use overall bias measures to adjust survey values, are the biases within a specified tolerance?

NASS currently conducts a redesigned validation survey for soybean OY. This survey is done in all states in the OY sample program. This design removed some unpopular aspects of the old validation surveys, including the concentration of work in one or two states and the variable workload resulting from changing states each year. Our goal was to verify the approximate 6% bias adjustment suggested by the historic series of studies. The current approach differs from prior studies. We now combine sources of error rather than trying to measure specific components. Thus, the results provide a basis for adjusting the survey for the many

small sources of error found in prior studies. These errors included: incorrectly measured row widths, field counts differing from lab counts, time lag bias due to the enumeration differing by several days from actual harvest, new planting patterns causing enumeration and imputation difficulties, enumerator fatigue errors, and plot location biases.

The rationale for the redesign begins with our estimator of state yield, the mean of the sample field yields, which is basically unbiased, except that we do not have the true field yield, Y ; but a sampled value, y . This estimate can be modeled as follows:

$$y_i = Y_i + B_{(f)i} + B_{(p)i} + B_{(ar)i} + B_{(u)i} + E_i$$

where

Y_i = True yield in field i .

$B_{(f)i}$ = The bias due to the within field estimator y_i , if any.

$B_{(p)i}$ = The bias due to field procedures.

$B_{(ar)i}$ = Partial nonresponse bias (for now the relatively small amount of complete nonresponse assumes the data is missing at random)

$B_{(u)i}$ = Other, maybe unknown, bias sources (editing, coding).

E_i = random error due to sampling two plots in the field.

We can measure the true yield for a subsample of the OY fields using yield weighted at a grain elevator, Y'_i . Thus,

$$Y'_i = Y_i + M_i$$

where

M_i = The "validation" measurement error at subsample field i .

We can compute for each field in the validated subsample (here again nonmeasurement of the validation value assumes data missing at random)

$$y_i - Y'_i = B_{(f)i} + B_{(p)i} + B_{(ar)i} + B_{(u)i} + E_i - M_i$$

making two reasonable assumptions:

(1) $E(E_i) = 0$, and (2) $M_i = 0$,

an estimate of total bias is

$$\text{Bias} = \Sigma (y_i - Y'_i) / n'$$

where n' is the validation subsample size.

Three years of data from the validation survey have produced the following results:

<u>Year</u>	<u>Estimated Bias in Bushels</u>	<u>Estimated Standard Error</u>	<u>Bias as percent of the Estimate</u>
1987	2.2	.9	5.8
1988	2.3	.8	7.6
1989	3.2	.9	8.7

Thus, the studies validated the 6% adjustment of the survey data as reasonable. Future research can determine the optimal use of the validation survey for adjustment. We also need to assess the implicit missing at random assumptions. We can get some ideas on the reasonableness of the assumption using farmers reported yields to measure group differences. We need the assumption that the biases measured by the validation survey are uncorrelated with the action of obtaining elevator yields. This assumption is reasonable, but should be tested occasionally. With the redesigned validation survey we have two of the three estimates (the OY yield estimate, the validation survey estimates of OY bias, and a nonresponse bias estimate). These are the estimates of the major error components which are necessary to assess the accuracy of the between-year of yield estimates.

Conclusion

Although the level of nonsampling error in establishment surveys was not directly measured in WP-15, nonuse of control and measurement techniques should not be interpreted as a lack of errors. Is it time for us to regain the balance between the importance which we put on the elements of survey quality and our actual practice? For too many years, emphasis in most government agencies has been on timeliness and resources (usually shrinking). It's time to shift more effort to relevance and accuracy issues. We might help ourselves by training users in survey quality concepts so they can help us prioritize our efforts and maybe lead the effort to secure more funding. Our easiest beginning in this road to quality could start merely by publishing more of what we do know about the errors.

Increased interest in organized quality efforts such as total quality management philosophies is promising. Organizations need to ask questions such as:

1. What measure(s) does top management use to quantify survey or organizational effectiveness? (Is it the same as the data users?)
2. How are these measures used to manage and plan for the long run?

Agencies need to assess their training needs. We will face at least some shortage of new hires with the survey research skills necessary. Some predict that the shortage will be acute and go beyond survey skills to general quantitative skills. Will agencies respond with creativity in developing staffing and training plans? We should do more to address this problem now.

Finally, WP-15, actually all the working papers, needs to be more widely read. (Only a small percentage of the audience at the presentation had seen WP-15.) Agencies and users can benefit by identifying errors which were not previously considered and/or techniques which could be used. I caution against being overwhelmed with the quantity of errors displayed in WP-15. Don't worry that you can't eliminate or measure them all at once. I doubt that you have all these errors. Yet, don't be complacent. To improve survey quality you need a strategy. The strategy should define a systematic approach to the improvement and measurement of the effects of existing error sources as well as proposed changes in the survey process. Be flexible as you move along with the strategy, enjoying small successes as they come and avoiding the expectation of overnight miracles.

References

Bailar, B.A., (1968) "Recent Research in Reinterview Procedures," JASA 63:41-63.

Fecso, Ron, (1986) "Sample Survey Quality: Issues and Examples from an Agricultural Survey," Proceedings of The Section on Survey Research Methods, American Statistical Association.

Fecso, R., R.D. Tortora and F.A. Vogel, "Sampling Frames for Agriculture in the United States," J. of Official Statistics, Vol. 2, No. 3, pp. 279-292, 1986.

Fecso, Ron and Brad Pafford, "Response Errors in Establishment Surveys with an Example From Agribusiness Survey," Proceedings of the Section on Survey Research Methods, ASA, 1988.

Francisco, C., W.A. Fuller and R. Fecso, "Statistical Properties of Crop Production Estimators," Survey Methodology Vol. 13, No. 1, June 1987, pp. 45-62.

Groves, Robert M. and Lou J. Magilavy, (1986) "Measuring and Explaining Interviewer Effects in Centralized Telephone Surveys," Public Opinion Quarterly, Vol. 50:251-266.

Matthews, R.V., "An Overview of the 1985 Corn, Cotton, Soybean, and Wheat Objective Yield Surveys," USDA, Stat. Rept. Ser., Staff Report. Nov. 1985.

Office of Management and Budget, Quality in Establishment Surveys, Statistical Policy Working Paper 15, Washington, D.C., 1988.

O'Muircheartaigh, Colm A., (1986) "Correlates of Reinterview Response Inconsistency in the Current Population Survey," Second Annual Research Conference, Bureau of the Census, March 23-26, 1985 in Reston, Va.

Pafford, Brad, (1988) "Use of Reinterview Techniques for Quality Assurance: The Measurement of Response Error in the Collection of December 1987 Quarterly Grain Stocks Data Using CATI," National Agricultural Statistics Service, Research Report, USDA.

Reiser, M., R. Fecso and K. Taylor, "A Nested Error Model for the Objective Yield Survey," Proc. of Section on Survey Research Methods, ASA, 1987.

United Nations, National Household Survey Capability Program, Nonsampling Errors in Household Surveys, New York, 1982.

DISCUSSION

David A. Binder
Statistics Canada

I would like to thank the organizers for inviting me as a discussant at this important session on Quality in Business Surveys. Prior to these meetings, I reviewed once again the Statistical Policy Working Paper 15, "Quality in Establishment Surveys", and I would highly recommend it be read by both novices and experienced survey statisticians who deal with the design or analysis of business surveys.

One clear fact which comes out of Working Paper 15 is that there are many issues and methods which are common to most federal business surveys. Certain issues faced in business surveys are more difficult than in social and demographic surveys. Part of this is due to the complex and dynamic structures within which the business community operates. When designing and conducting such surveys, it is important to keep in mind the operational realities of the business world.

Since there are many commonalities among business surveys, statistical agencies should pool their knowledge and expertise to take advantage of their combined experience. For example, there are sufficiently many common practices for sampling, data collection, editing, estimation and dissemination of the results, that certain standards and guidelines could be developed among the agencies. Sharing information and expertise is a worthwhile objective which meetings such as this can help accomplish. Whereas legalities of data sharing pose some obstacles at present, hopefully these can be overcome in the longer term.

There are, of course, many aspects to improving the quality of business surveys, including frame issues and non-sampling errors. The development of general purpose business frames can lead to sophisticated and expensive systems, especially with respect to development and maintenance. This is because a general purpose frame should reflect the realities of the operating structures in the business world and there must also be user-friendly interfaces with such a frame. In practice, there is often a gap between conceptual frameworks and actual application.

Quality of the Frame

An important area of concern in the quality of business surveys is the quality of the frame itself. Survey quality will depend on the quality of the frame information as well as the ease of accessibility to the frame data. Frames can never be perfect. Some of the sources of error are:

- undercoverage, especially for births
- overcoverage, especially due to duplication and inclusion of out-of-scope units
- misclassification of industry code, employment size, other size measures, etc.
- identification of appropriate reporting units (collection entities) which reflects the operating structure of the business

It is important to include in the development of a frame a Program to measure the quality of the frame information. This is particularly true when the frame will be used by a variety of users other than the developers themselves. Examples of quality measures are:

- size of the backlog for SIC classification
- distribution of lag times for births and other updates to the frame
- errors resulting from cutoffs for multi-unit employers
- duplication
- matching errors

If the frame is to contain the most up to date information, there should be some facility for incorporating and verifying feedback from the surveys themselves. This can lead to complications, where the information being derived from one survey may affect other surveys (e.g. a change in the relationships among multi-unit employers).

Structure of the Frame

If it is anticipated that the Business Establishment Listing (BEL) of the Bureau of Labor Statistics will be used by other agencies conducting business surveys, it should be noted that many of their needs cannot be met within the framework being discussed here. The administrative world does not always correspond to the business world. A listing which is useful for employment and related labor characteristics may not be suitable for surveys of economic production and other special characteristics.

The structure of the BEL for multi-unit employers needs some clarification. Whereas the worksite may be able to report employment data, it may not be able to report on profit and loss or balance sheet data. Different reporting units (collection entities) may need to be identified for different surveys. It

cannot be assumed that the respondent will necessarily conform to your concepts.

At Statistics Canada, we have developed a hierarchical structure of statistical entities for the larger businesses. These are (i) the enterprise, where a full set of consolidated financial statements are available, (ii) the company which can report on profit and loss and other balance sheet items, (iii) the establishment, which can report on such items as value of output, cost of intermediate inputs, inventories, number of employees, and salaries and wages, (iv) the location, which can report sales and number of employees. This recognizes the relationship between the business world and the statistical needs for economic surveys. However, it is a complex structure to maintain.

Retrieval Systems

Not only are frame maintenance procedures resource intensive, but effective retrieval systems can be quite complex and expensive to develop. Quality improvements to business surveys through better quality frames can only be realized if the frame information is easily obtained both cross-sectionally and through time. Examples of some of the needs which are expressed by users of frame information are:

- linking of data through time
- historical files
- response histories
- linking of data within enterprises
- identification of seasonal and volatile firms
- having sufficient structure to roll up to enterprise and track changes in structure over time
- survey feedback (and verification)
- requirements for estimation (regression, ratio, composite, benchmarking, poststratification)

Other Frame Considerations

The needs of the frame will change depending upon the survey frequency and the reference periods. For example, the units considered in-scope could vary according to whether the survey is monthly, quarterly or annual.

Even with all the complexities I have mentioned regarding the development and maintenance of business frames, I would strongly encourage such development, with any deficiencies explicitly laid out. One of the uses of a high quality frame is the ability to perform analyses of business demographics, showing behaviour of births, deaths, mergers and amalgamations, which is an important side benefit.

Total Survey Error

As was pointed during the session, improving frame quality is only one of the many mechanisms to meet the overall objective of controlling survey errors. Development of survey quality profiles has been mentioned as an important tool to monitor, control and manage surveys.

Response errors should be a particularly important concern to the survey-taker. However, response errors are often due to the survey instrument itself, rather than the respondent. Recent experiences with cognitive methods have proven useful here. Often there are trade-offs between ideal concepts and the respondents' ability to respond accurately. For example, when asking a farm operator about value of equipment on land which he operates, he may prefer to report on equipment which he owns but which may be situated on another farm, rather than including equipment which is owned by someone else, but which is situated on his land. This creates difficulties for the survey-taker who is trying to avoid coverage errors. These are not easy problems to overcome, but the first step in all these endeavors is to recognize the problem and possibly measure its impact. Without special studies, it would be difficult to assess the relative merits of coverage error on the one hand and response error on the other.

In general, we need to concentrate on methods to synthesize all the errors into an overall measure of survey quality. This would allow informed decisions to be made regarding the relative merits of improving one survey process over another. If such a model existed, we could answer some common concerns such as the relative contribution of edit and imputation to the reduction in total survey error and whether simpler methods could achieve comparable results.

One possibility would be to use develop a microdata simulation database which incorporate as many of the known errors as possible. This database would consist of microdata which look like the real population. Various models for response and nonresponse errors could be simulated and then the data would be processed using existing or proposed methods. Since the original "true" data are known, we could assess the relative impacts of improving survey coverage versus using an alternative estimator versus adding more edits to the survey process, for example.

DISCUSSION

Charles D. Cowan
Opinion Research Corporation

What These Papers Have in Common

If there is a single message that comes through in both the papers being discussed, it is that:

Avoidance and/or Control is the Best Approach in Dealing with Nonsampling Error.

Quality is something that one builds into surveys and continues to monitor. While one cannot completely avoid problems in surveys, it is markedly better to avoid or control a problem than it is to attempt to make an a posteriori correction to fix the problem. Such a fix usually is based on a much smaller amount of information collected from a supplemental sample or survey and adds variance to the original survey estimates. It is also usually the case that a fix introduced at the end of a survey only takes care of one problem and is not very cost efficient.

In their paper, Tupek and MacDonald describe a process of expanding a sampling frame for business surveys that addresses several different sources of nonsampling error. Their work with the sampling frame deals with coverage issues, timing issues, definitional problems in the surveys, estimation, use of administrative records for weighting and variance reduction, and other aspects of the conduct of business surveys. Their approach is to improve the basic materials used for surveys to encourage more efficiency and accuracy at later stages.

Fecso in his paper describes a process of measuring and controlling as many aspects as possible of incidence of nonsampling error. He also supports the idea that nonsampling error is best dealt with by avoidance, but is also realistic in suggesting that a catalog of problems is useful for two primary purposes: planning future surveys and providing documentation for users of the current effort. This control process can be used to ensure that the data produced in a survey are of the best quality given the constraint that control is imposed as part of the process, since many types of nonsampling errors cannot be totally avoided.

Specific Quality Issues for Business and Establishment Surveys

As one reads and compares these papers, one is reminded of the fact that business and establishment surveys are different household surveys in several key ways:

- 1) The availability of attributes on the frame and the use of this frame information at the unit level differs from what can be done in household surveys,
- 2) The surveys themselves make extensive use of records as a basis for reporting, and
- 3) The data to be collected in business and establishment surveys has a multilevel nature, meaning that information about the businesses is hierarchical and we are interested in the information at each level (e.g., Sears Headquarters, regional offices, distribution centers, and individual stores).

These factors are crucial to the design of business and establishment surveys. Use of information on the frame for design and use of records in collection makes it possible to improve the quality of these types of surveys relative to household surveys, but this is counterbalanced to an extent by the complications introduced by the multilevel nature of the data to be collected.

Tupek and MacDonald note in their paper that for the surveys they conduct that establishments come from skewed populations, and having this information on the frame makes it possible to design a survey that is much more efficient, especially for multiple characteristics to be measured simultaneously. However, reliance on this information in the frame makes the accuracy of frame information crucial at the individual unit level for both sampling and estimation purposes. Their project on frame expansion and improvements has an impact in several areas. The first is sample frame development, so that more business and establishments are represented. This is broader than a coverage issue, since coverage is usually viewed as a problem that pervades an extant frame. Tupek and MacDonald address coverage issues in this way, but also include whole segments of the business population previously excluded from the frame.

A second area impacted by the frame expansion and improvements project on which they report is the actual design of the sample, where the sample can be optimized for making different types of estimates using information available on the frame. A third area impacted by the frame expansion and improvements is in data collection, and the final area is in estimation. Tupek and MacDonald point out that the new frame encourages the conduct of new longitudinal surveys, the selection of sample at the unit of analysis (instead of collecting the information by proxy or sampling down to the unit of analysis after starting at a higher level in the hierarchy), improvement in response rates because of higher eligibility rates, savings in terms of time and effort expended on the survey, and improvement in weighting and ratio estimation procedures.

Fecso takes a different approach to dealing with nonsampling error. He catalogs sources of nonsampling error, and his approach is to detect, measure, and control the nonsampling error. Many of the sources of nonsampling error he lists are common to both household and business surveys, but with business surveys he has a variety of records, including past survey collections, available for detection and measurement of nonsampling error.

A primary concern for the use of records is the accuracy of the data in the records, since the records themselves could be in error. Although not mentioned in the paper, some of the most interesting work in health care surveys is modeling of nonsampling error when hospital records and information based on patient recall don't match and either is potentially wrong. The same is true for business surveys -- accuracy in the records systems is crucial for detection and measurement of nonsampling error as part of a quality management system for a survey. Another factor related to accuracy is the consistency of definitions used by different respondents. If the data are accurate but based on different definitions, then there is a problem in how the data might be used for detection and measurement of nonsampling error.

Concerns with Business and Establishment Surveys Not Covered

While both papers are excellent in the way they cover in depth quality issues facing business and establishment surveys, they both miss some salient points peculiar to these types of surveys. The first was mentioned earlier, namely that businesses are hierarchical, which leads to some difficult questions regarding who reports in these surveys, and how the various businesses relate to one another (i.e., at what level do we define the unit of analysis?). In terms of how units relate, an example was given earlier for Sears, which owns not only Sears Retail, but also has Allstate Insurance, a mailing service, regional offices, catalog stores, and local retail stores. Are we interested in these surveys in getting reports from the lowest level in this chain? How does Sears headquarters report exactly -- for itself as an establishment with a certain number of employees, or does it include all employees and sales at all locations? If there is confusion in reporting rules for a survey, we could wind up with severe overcounting or undercounting of activities and personnel.

Another issue has to do with the reporting of activities within a firm. In reporting mailing activities, for example, each firm and each location of a firm will have some activities to report. To whom do we speak in the firm to get a complete picture? There are separate operating units within firms, each with a manager knowledgeable about his own unit's activities. And there are sometimes other units that assist in terms of technical or operational support. Do we talk to managers in both or all offices

or units, or is there a central source that can answer all questions knowledgeably and without duplication?

There are two final concerns we have regarding quality in business and establishment surveys. One has to do with the process of improving and expanding the frame for a business survey, which usually translates into adding smaller firms. These firms are more likely to be related to other members of the population, and they are more prone to movement in and out of the population (births and deaths). Because of these factors, they add a certain amount of instability to the estimation process. This may be good or bad -- on the one hand we have a more realistic representation of the population of businesses when we include more firms, but on the other hand for certain types of statistics we may be adding more variation without a real gain in forecasting or descriptive accuracy. This problem could be labeled: "messiness at the edge".

The other problem not addressed in either paper, and of particular concern in the Fecso paper, is that a large, well conceived and executed survey might not benefit from a Nonresponse/Nonsampling Error Correction that is estimated from a small onetime experiment. While in theory the idea of implementing research studies to monitor the quality of ongoing surveys is laudable and should enhance the quality of the surveys, implementation for Federal surveys often falls a bit short, with a simple, one-time study implemented to measure a particular problem. A small scale, high variance research study should be viewed as just that, and not a vehicle for making corrections to a multimillion dollar effort. If the nonsampling error problem is sufficient to justify such an effort, and the nonsampling error cannot be dealt with as part of the design, then sufficient resources should be devoted to measurement and control to take care of the problem. Essentially, the problem becomes one of design again, with focus on the proper allocation of resources between the survey and the experiment to fix the survey.

Conclusions

Both papers were excellent summaries of the state of the art for measuring and maintaining quality in Federal surveys of businesses and establishments. Researchers involved in the design of either business or household surveys would benefit from studying and implementing the principles found in either paper.

Session 8
COGNITIVE LABORATORIES

THE BUREAU OF LABOR STATISTICS' COLLECTION PROCEDURES
RESEARCH LABORATORY: ACCOMPLISHMENTS AND FUTURE DIRECTIONS

Cathryn S. Dippo
Douglas Herrmann
U. S. Bureau of Labor Statistics

I. Introduction

The accomplishments of the Cognitive Aspects of Survey Methodology movement (Jabine, et al. 1984) have clearly been substantial. This is especially true in Washington, where three Federal agencies (Bureau of the Census, Bureau of Labor Statistics (BLS), and the National Center for Health Statistics) have established laboratories.

Consider the scope of BLS' survey research programs. Most of the sampling units from which data are collected by or for BLS are establishments. While approximately 60,000 households are questioned about labor force participation each month in the Current Population Survey (CPS), 340,000 establishments are being asked to report their payroll employment each month in the Current Employment Statistics Survey. More than 200,000 price quotes are being collected each month from establishments in the Consumer Price, Producer Price, and International Price Index programs. Moreover, much of the data are currently being collected by mail, without person-to-person interaction. In the future, more and more of the data will be collected with computer assistance, and the human-machine interface will take on added importance. Furthermore, in most establishment surveys, the needed data can be directly observed (e.g., consumer prices) or exist in records rather than in the memories of the respondents. Even in household surveys, many respondents are being asked to recall not only autobiographical events, but also information that exists in household records and information about other members of their household.

Thus, the mission of the Bureau requires the BLS laboratory to consider more than just questionnaires to be used with personal visit interviewing in the context of a household survey about autobiographical events. The Bureau acknowledged this fact when selecting the name for its laboratory -- the Collection Procedures Research Laboratory (CPRL) -- which was established in 1988. The basic goal of the CPRL is to improve through interdisciplinary research the quality of data collected and published by BLS. As originally envisioned, all forms of oral and written communication used in the collection and processing of survey data are appropriate subjects for investigation, as are all aspects of data collection, including mode, manuals, and interviewer training.

The CPRL's staff includes cognitive psychologists, social psychologists, sociologists, and a psychological anthropologist. For most of their projects, they work closely with the economists or program specialists responsible for defining the concepts to be measured by the Bureau's survey programs. To augment staff resources, the CPRL has labor hour contracts with the Institute for Social Research at the University of Michigan and Westat, Inc. The laboratory also does work under contract for other Federal agencies such as the Internal Revenue Service.

Although the CPRL has only existed for two years, its research program has been both broad and prolific. In section II, some accomplishments of the CPRL are reviewed. The discussion is organized within the framework of an information processing model. In section III, some directions for future research are described. The success of focusing on the cognitive system suggests that focusing on other behavioral systems may produce further gains in data quality through improved survey theory and practice. Moreover, the success of using laboratory techniques for investigating the data collection processes used in sample surveys leads us to believe the techniques can be useful in improving other aspects of survey design.

II. Accomplishments to date

The CPRL has integrated the cognitive approach into the Bureau's survey research program to good effect in many ways. Primarily, the laboratory has changed how data collection research is conducted at BLS. Not only has the research conducted to date affected our understanding of the survey process, but the fact of its existence has heightened awareness throughout BLS of the need for a better understanding of all aspects of the data collection process (Norwood and Dipbo in press).

Some results of the CPRL's research efforts are presented here within the framework of an information processing model (Cannell, et al. 1989; Tourangeau 1984) that has four distinct stages: comprehension, retrieval, judgment, and communication. As applied to respondents, these stages refer to the comprehension of a question, retrieval of pertinent information, judgment about the accuracy of the information retrieved, and communication about this information within social and other restrictions imposed by the survey situation. As applied to interviewers, these stages may refer to comprehension of the question, retrieval of appropriate ways to say the question aloud, judgment about whether the respondent has understood the question, and communication to ensure the question has been understood (such as by rereading it) or, if the question has apparently been understood, to indicate that another question is about to be presented.

A. Comprehension

Question comprehension clearly requires that the terms making up a question be correctly understood. The accuracy of term comprehension has been shown by many psycholinguistic investigations to differ in certain ways.

Multiple meanings of terms: A term may lead some respondents to answer inappropriately because it may convey a meaning different from that intended by the designer. Research at BLS has accordingly attempted to identify terms with several meanings that are not made explicit by the phrasing of questions and might be likely to produce misinterpretations. Since the issue of employment is of personal significance to most people, questions about employment status are likely to predispose respondents (especially the unemployed or those with insecure employment) to be influenced by social desirability when answering the CPS (DeMaio 1984; Edwards, Levine, and Allen 1989). The misinterpretation of employment status terms may easily occur in a survey such as the CPS (Martin 1987).

Accordingly, respondents' interpretations of two key terms on the CPS concerning unemployment status, "on layoff" and "looking for work," have been examined. The CPS definition of unemployment refers to persons who were not employed during the survey week, were available for work, and had made specific efforts to find employment sometime during the prior four weeks. Persons who are waiting to be recalled to a job from which they have been laid off need not be looking for work to be classified as unemployed. As expected, research demonstrates that these terms are sometimes misinterpreted by laboratory respondents to the CPS. Similar research into the effects of multiple meanings of terms has also been conducted for several sections of the Consumer Expenditure (CE) Interview Survey, including the sections on medical care, home purchase, and trip expenditures (Miller and Downes-LeGuin 1989). Since our results indicated that people interpret "payments" in different ways, the section on medical care expenditures has since been modified to avoid misinterpretations of this term.

Diverse meanings: Diversity of term meaning also may impair comprehension. For example, in a recent pilot survey of business establishments, respondents were asked to report all "nonwage cash payments" paid to employees during the calendar year. BLS defined the payments to include bonuses and awards, lump-sum, cash profit sharing, and severance payments, and nonregular commissions, but since this technical term probably was not too familiar to respondents, the meanings of "nonwage cash payments" can be expected to vary across respondents. When the interpretations of this term by respondents were investigated, it was found that respondents interpreted "nonwage cash payments" in a diverse fashion. Some interpreted it too broadly to include payments in kind, such as a new car (Boehm 1988), and some too narrowly to

include only cash and not cashable checks (Phipps 1990). Another group of respondents who had made such payments simply checked they had made no payments because of a lack of understanding of what the term included. Respondent exclusion and nonreporting of payments were more serious comprehension errors than inclusion of inappropriate payments, contributing to underreporting.

Format properties: When respondents complete a survey form received in the mail, the format of the instrument may play a crucial role in the respondents' comprehension. If the format does not make it clear what parts of the instructions are essential, respondents may overlook these parts and respond inappropriately. For example, in the Nonwage Cash Payments Pilot Survey (Phipps 1990), instructions, definitions, and examples were on the back of a one-page questionnaire, for which two different layouts were used. One layout required respondents first to provide an annual nonwage cash payment total and an annual payroll total, then answer a set of yes/no questions asking if they made specific types of nonwage cash payments. The second layout placed the set of yes/no questions first, with the payments and payroll totals requested at the bottom of the page. Reporters receiving the second layout were much less likely to provide the annual payroll total, stating in retrospective interviews that they overlooked it or did not understand they were to provide it. Thus, the layout of the second form, combined with a lack of instruction, caused an entire section of the form to be overlooked. As expected, the format of a survey played an important role in the respondents' comprehension of survey items.

The types of cues used on a self-administered form like an expenditure diary also can affect comprehension. In developing a diary for recording clothing expenditures, alternative cueing levels were tested in a laboratory. Results indicated that a shorter diary with multiple pages that repeated the general cues, e.g., buying clothes, was more effective than a longer, more structured version with specific cues. Respondents were better at clarifying the domain of purchases to be recorded with the general cues than with the specific cues, i.e., the specific cues led them to restrict their comprehension of listed items more narrowly than intended.

B. Retrieval

Most Federal surveys require respondents to retrieve information about factual or autobiographical events. Faced with the need to control data collection costs, the time period for which the events are to be recalled is often long. For example, the reference period for the CE Interview Survey is three months. In the CPS, respondents may be asked questions about last week, the last four weeks, or the last time they worked, which could require recall for a long period of time. (For further discussion of

memory retrieval errors in CE and CPS, see Dipbo 1989 and Mullin 1990).

Cues: Often a situation is inadequate in the cues it presents for retrieval. Alternatively, when enough appropriate cues are brought forth, a person can retrieve the previously "forgotten" memory. While some information is probably lost from memory due to diseases and environmental influences (such as alcohol), cues clearly play an important role in retrieval. Accordingly, several investigations have attempted to increase response accuracy on surveys by providing additional cues to retrieval, e.g., Lessler, et al. (1989). Still, it is important to recognize that some cues can be misleading and ensure that a respondent does not retrieve the appropriate information. Cues facilitate only when they correctly direct retrieval.

In the Nonwage Cash Payments Pilot Survey, underreporting was investigated by presenting cues to facilitate retrieval. When respondents (company representatives) were given specific cues pertaining to bonus and award payments, recall of such payments was 11 percent higher than without cues (Phipps 1990). Also, in the CE Diary Survey, cues with varying levels of generality have been tested. For example, general cues included "beef (ground, roasts, steaks, briskets, etc.)" and specific cues included "ground beef, chuck roast, round roast, other roast, round steak, sirloin steak, other steak, other beef and veal." Underreporting was greater with general cues for certain items, particularly nonfood items. On the other hand, the level of reporting for many food items was not affected by the type of cues (Tucker and Bennett 1988).

Strategies: To get accurate recall about the past, it is necessary to get people to retrieve the mental records of what they actually did. Several strategies to get respondents to access their memories of experiences have proved useful in our investigations at BLS. One strategy has respondents recall a critical personal event that occurred in the reference period in order to anchor the period. A second strategy has a respondent consult a calendar when attempting to recall. A third strategy has respondents decompose events recalled into smaller events to ensure that what is being recalled is a real experience and not a stereotypical schema. Research funded by BLS has found that respondents vary in the extent to which they employ the strategy that they were instructed to use. Only one-third of the laboratory subjects instructed to use a decomposition strategy when responding to questions on their hours worked used the strategy. Also, the vast majority of proxy respondents presented with this strategy ignored it because they did not have the knowledge necessary to use it.

Expertise: In a laboratory study of household respondent pairs using the CPS questionnaire, proxy responses disagreed with those of the self-respondent approximately one-third of the time

(Boehm 1989). In another laboratory study, when respondents were instructed to use the decomposition procedure, the vast majority of proxy respondents ignored the procedure, since they did not have the knowledge necessary to use it (Edwards, et al. 1989). Self-respondents were found to overreport and proxy respondents to underreport the hours worked. Also, proxy respondents were more likely than self-respondents to make errors, and their errors tended to be larger (see also Tanur 1990). As might be expected, proxies fail in areas they are less likely to know about. For example, proxies underreport more when the person reported on worked weekends or worked extra hours. Also, proxy error was greater when the respondent was unrelated to or from a different generation than the person to whom the data related (Edwards, et al. 1989).

C. Judgment

People may recall correctly but not realize the recalled information is correct. They may recall correct information, know it is correct, but express it inappropriately because they misconceive how responses are to be expressed. It was noted above that field research on the CE Diary Survey indicated specific cues were often more effective and led to less underreporting than general cues (Tucker and Bennett 1988). Laboratory research has indicated that judgment is also a factor. When given specific cues, laboratory subjects were sometimes unsure of where to record products on the form. Whether this hinders reporting is still an open question, but the accuracy of reports is affected (Tucker, et al. 1989). The specific cues also may make the task more onerous.

D. Communication

The importance of communication to cognition has largely been recognized in social psychology and anthropology. A considerable amount of survey research has shown that respondents' inclination to answer questions may be affected by the social desirability of the answers. In some cases, respondents may be disinclined to answer because they do not want to share certain kinds of information. In other cases, they may not want to present themselves in a bad light. In other cases yet, they may want to adapt their response to what they perceive to be the expectations of the interviewer.

While BLS has yet to complete an investigation of communication, it has recently begun several such investigations. First, the laboratory is conducting research into the psycholinguistic factors that persuade a respondent to provide confidential information to a survey (Herrmann, et al. 1990). This research will indicate the degree of trust elicited by different protection terms (confidential, private, secret, concealed,

nondisclosed). Second, we are examining the influence of interviewer errors on the errors of respondents using techniques developed by Cannell (Cannell, et al. 1989). For example, tape recordings of CE Survey interviews are being analyzed to determine whether the quality of answers produced by respondents varies with the quality of the interviewers' presentation of a question. Third, like other agencies, we are investigating the use of computer-assisted telephone interviewing (CATI) for some BLS surveys. Research is underway for the CPS, CPI-Housing, and Continuing Point-of-Purchase surveys to determine if people respond in the same manner in a computer-assisted telephone interview as they do in a personal interview. It has been suggested that the personal interview ensures better attention from the respondent, but it has also been suggested that CATI elicits information that otherwise might not be disclosed because the respondent feels less personally involved when interacting with an interviewer on the telephone. In various ways our research is addressing these alternative expectations about CATI.

III. Future directions

Prior to the establishment of the laboratory, BLS sponsored a Questionnaire Design Advisory Conference to seek advice on the types of questionnaire research that should be undertaken for the CE and CPS (Bienias, et al. 1987). The conference participants all advocated the incorporation of cognitive concepts into the BLS research program and suggested that research focus on the issues of respondent rules, respondent and interviewer roles, questionnaire form and content, and statistical estimation.

In addition, our ongoing research program has taught us that many aspects of the data collection process require a broader integrated-systems approach rather than a cognitive approach to research. The accuracy and efficiency of survey responses are affected not only by cognitive variables (e.g., abstractness of terms, retrieval cues) but also by other kinds of variables (e.g., physiological, perceptual, emotional, motivational, social, societal, cultural, and economic; see Royce 1973). In some cases, these variables affect responding because they interact with the quality of cognitive processes underlying responding. In other cases, these other variables leave cognitions unaffected but instead interact with a respondent's inclination to report accurately about these cognitions.

A. Looking beyond the cognitive approach

An integrated-systems conception of cognition has been advocated increasingly in recent years by scholars in anthropology (Cole and Scribner 1974), psychology, and neuroscience. Some noncognitive psychological and societal factors that may affect the

response process are: Physiological condition, perception, emotional state, motivation, familial roles, and societal norms.

Physiological condition: The accuracy and efficiency of cognitive responses are affected by the physical state of a person's body (Squire 1987). Physiological condition, as affected by physical health, influences a person's ability to understand, remember, reason, and analyze. A variety of routine health conditions (such as the common cold) may impair the accuracy and/or efficiency of cognitive processes (Cutler and Grams 1988). Cognitive processes are also impaired by commonly imbibed substances, such as coffee, tobacco, tranquilizers and antidepressants, and even certain antibiotics.

The CPRL has been sponsoring laboratory research on the effects of computer-assisted personal interviewing (CAPI) on the interviewer (Couper, et al. 1990). Although the studies have been within the context of the Consumer Price Index survey, where interviewers conduct interviews both on the doorstep of housing units and walking the aisles in retail establishments, the procedures developed, concerns raised, and results are generally applicable. For example, more than 40 percent of the 46 interviewers who volunteered to be laboratory subjects stated that they had suffered neck, shoulder, and/or lower back problems in the 12 months prior to any contact with a portable computer. Moreover, approximately 75 percent of the subjects wore some form of corrective lenses, with bifocals presenting particular problems for interviewers trying to focus on the keyboard, screen, and respondent.

Perception: The quality of visual stimuli affects the ease of reading and comprehension. The role of perception is of special importance in many Federal surveys where data are collected via a self-administered form. For these surveys, the perceptual constructs may have significant effects on the quality of data. Wright (1980) suggests classifying form-design issues into three categories: the language of forms, overall structure, and the substructures within the forms such as the questions themselves. In addition, there are perceptual issues related to the appearance of questionnaires, such as color and print font.

The presence of visual stimuli affects retrieval processes more than thinking about or imagining the stimulus. For example, psychological research indicates that the frequency at which academics use external aids, such as files and piles of papers on one's desk, has been found to be positively correlated with scholarly productivity (Hertel 1988). Survey research indicates that expenditure reporting increases with the use by respondents of an information booklet describing the types of items that belong to the categories being read aloud by the interviewer. More respondents appear to be willing to read the item lists than to listen to an interviewer read the list to them.

Respondents to the Occupational Safety and Health Survey face a very difficult task in deciding if an incident is an injury or an illness and if it is reportable or not. Currently, respondents receive a 22-page set of guidelines. Laboratory staff are now investigating different methods for communicating the decision logic to respondents, i.e., flow charts or graphic representations of the decision paths. In addition, a simple user's guide (no more than 10 pages) is being prepared for respondents who are new to OSHA recordkeeping. Unlike the longer guidelines, this guide contains background on the 1970 OSHA act and provides examples on how to recognize, record, and report occupational injuries and illnesses.

Emotional state: Our cognitive ability to comprehend, retrieve, evaluate, and respond may be affected by our emotional state (Wolkowitz and Weingartner 1988), which in turn may be affected by recent events or prolonged stress. Stress, a major factor moderating emotional states, has been associated with cognitive failures in everyday life. Sometimes, emotional states may prevent people from producing correct responses that they "know" at some level. For example, despite decades of controversy, it is now generally accepted that sometimes people repress memories.

Nontrivial levels of stress are currently experienced by interviewers. With the change over the next decade to increased CATI, the possibility of increased interviewer stress is real. In surveys like the CPS, the proportion of personal visit interviews will increase for most interviewers working in large metropolitan areas as many of their telephone interviews are transferred to a centralized CATI facility. Concerns about personal safety and administrative pressures to maintain high response rates are but two factors which may contribute to increased interviewer stress. In a centralized CATI facility, interviewers know their work is constantly being monitored. Recent news stories about the effects of constant observation and work quotas in the telephone industry indicate stress levels can be very high in these kinds of situations.

Motivation: We know little about respondents' motivations for responding to survey questionnaires. Census' recent experience of overestimating the mail-return rate in the decennial census is but one indicator of how little we know. At BLS, those of us working on the CE Interview Survey constantly wonder why anyone would agree to an interview that is expected to last 2 hours. To investigate survey respondent motivation, a large-scale research project on household survey response has been initiated by Robert Groves at the University of Michigan, sponsored by the Bureau of Justice Statistics, the Bureau of Labor Statistics, and the National Center for Health Statistics. One part of the project is an examination of both interviewer (e.g., attitudes, behavior, and characteristics) and administrative (e.g., procedures, workload

levels, design parameters) influences on survey participation (Groves, R.M. and Cialdini, R. 1990). To examine the effects of alternative forms of persuasive communication on sample attrition rates and item response rates, BLS is conducting experiments using appeals that stress the use of Current Employment Statistics data by the trade associations representing the establishments (McKay 1990).

Familial roles: The roles people assume within the family have been found in recent years to affect cognitive processes. While it may be assumed in some surveys that people within a home are equally able to answer questions pertaining to the household, research shows that different family roles carry responsibility for knowing about certain kinds of information. For example, wives tend to know more about the health and activities of children, whereas husbands tend to know more about how community activities affect the household. Single parents tend to know the information possessed by both spouses in dual-parent households.

With the prevalence of proxy reporting in most household surveys, the importance of learning about what information is exchanged within households and how should not be understated. Recent research on proxy reporting in the CPS indicates adults may be worse proxy reporters for youths than for other adults in a household (Tanur 1990). Moreover, the proxy reporting of job search may be dependent upon the type of job search strategies being used by youth. As Tanur notes, there is no literature about family communication patterns and the issue of who in the family talks to whom about what.

Societal norms: Cognitive performance is affected by groups in several ways. For example, people are disinclined to perform memory tasks when the social stereotypes that apply to them indicate that they cannot perform well, such as the stereotypes associated with age or with gender. Also, people will sometimes knowingly give the wrong answer to a question because they recognize that their answer is contradicted by the other members of a group.

Moreover, social pressures sometimes dispose people to communicate falsely what they do or do not know in order to achieve social goals. For example, people may say they cannot recall some event or information to avoid or speed up the questioning or to make a certain impression on the questioner. We do know that social desirability plays a role, but there has been little research into understanding the role (DeMaio 1984). We also know that the mode of data collection appears to have an effect on data, but we do not know why (Shoemaker, et al. 1989). Recent research by Suchman and Jordan (1990) shows clearly the influence of social and cultural variables.

Evidence indicates that members of all cultures can equally perform all manner of cognitive tasks if the environment has provided the cultures equivalent education and experience. However, because cultures typically involve different educational systems, belief systems, and occupational opportunities, members of different cultures acquire different cognitive skills (Cole and Scribner 1974). Thus, members of different subcultures of a multicultural society will interpret certain concepts differently and answer differently.

B. Looking beyond the interviewing process

The research laboratory and laboratory techniques can be used in a variety of survey design applications. Just as the responding process is affected by noncognitive variables, the survey process consists of more than just question answering. The entire survey design process, from defining the concepts to be measured through analyzing the data, involves the communication of concepts between people with different knowledge bases or an interaction between people and things. The process can benefit from a broad range of interdisciplinary research including both cognitive and other areas of psychology, other behavioral sciences, and human neuroscience.

The importance of the role of the interviewer has long been recognized. Data collection and training methods designed to control interviewer error, such as structured questionnaires and verbatim training, have been developed in an attempt to control interviewer error. Interviewer training typically stresses the need for neutrality, the use of specified questionnaire wording and administration procedures, and appropriate probing techniques. Recognizing the importance of this source of error, many BLS-sponsored laboratory studies conducted in the last two years have focused on the interviewer. These studies indicate the role of the interviewer can be studied effectively with laboratory techniques. Thus, it seems natural to expand our research in this area.

IV. Summary

As survey researchers, we really know very little about the psychological processes underlying interviewer and respondent behavior. The few laboratory studies to date indicate the cognitive approach is very useful. With this approach we are learning about the roles of comprehension, recall, judgment, and communication in the survey response process. Eventually, as we learn more, we can develop detailed models which questionnaire designers can use to assess new questions and forms for survey data collection.

Just as the research to date has shown that the cognitive approach is effective, it has shown that a more broad-based approach is necessary. Survey responses clearly emanate from all

behavioral systems within and outside the respondent. An understanding of how responding is affected by the cognitive system is not enough. A respondent's behavior is influenced by physiological, emotional, social, societal, and economic variables. A complete explanation of responding requires an understanding of all systems and how their influences are integrated overall to produce a response.

The adoption of an integrated-systems approach would be a natural step in the evolution of survey science. Consider the disciplinary history of economic statistics. First, there were economists producing simple descriptive statistics. The discipline of mathematical statistics was not really incorporated until probability sampling became the basis for sample designs. Then came the advent of computers. Just as we have expanded our use of statistical theory as applied to survey research beyond just sampling (e.g., to incorporating operations research techniques in sample design optimization and iterative methods such as raking in survey estimation), survey research may progress further by making use of not only cognitive psychology but also of knowledge of other psychological and sociopsychological systems.

References

Bienias, J., Dipppo, C., and Palmisano, M. (1987), *Questionnaire Design: Report on the 1987 BLS Advisory Conference*, Washington, DC: U.S. Department of Labor, Bureau of Labor Statistics.

Boehm, L. (1988), "CES Nonwage Cash Payment Prepilot Interviews," Internal memorandum to Alan Tupek dated December 16, Washington, DC: U.S. Department of Labor, Bureau of Labor Statistics.

Boehm, L. (1989), "The Relationship Between Confidence, Knowledge, and Performance in the Current Population Survey," in *Proceedings of the Section on Survey Research Methods*, American Statistical Association, in press.

Cannell, C., Fowler, F., Kalton, G., Oksenberg, L., and Bischooping, K. (1989), "New Quantitative Techniques for Pretesting Survey Questions," in *Bulletin of the International Statistical Institute*, pp. 481-495.

Cole, M. and Scribner, S. (1974), *Culture and Thought: A Psychological Introduction*, New York: John Wiley and Sons.

Couper, M., Groves, R., and Jacobs, C. (1990, in press), "Building Predictive Models of CAPI Acceptance in a Field Interviewing Staff," in *Proceedings of the 1990 Annual Research Conference*, Washington, DC: U.S. Department of Commerce, Bureau of the Census.

Cutler, S.J. and Grams, A.E. (1988), "Correlates of Self-Reported Everyday Memory Problems," *Journal of Gerontology*, 43, 582-590.

DeMaio, T. (1984), "Social Desirability and Survey Measurement: A Review," in *Surveying Subjective Phenomena*, eds. C. Turner and E. Martin, New York: Russell Sage.

Dippo, C.S. (1989), "The Use of Cognitive Laboratory Techniques for Investigating Memory Retrieval Errors in Retrospective Surveys," in *Bulletin of the International Statistical Institute*, Vol. LIII, Book 2, pp. 363-382.

Edwards, S., Levine, R., and Allen, B. (1989), "Cognitive Strategies for Reporting Hours Worked," in *Proceedings of the Section on Survey Research Methods*, American Statistical Association, in press.

Groves, R.M. and Cialdini, R. (1990), "Toward a Useful Theory of Survey Participation," unpublished manuscript.

Herrmann, D., van Melis-Wright, M., and Stone, D. (1990), "The Semantic Basis of Confidentiality," in *Proceedings of the Section on Survey Methods Research*, American Statistical Association, to appear.

Hertel, P. (1988), "External Memory," in M. Gruneberg, P. Morris, and R. Sykes (eds.), *Practical Aspects of Memory*, New York: John Wiley and Sons.

Jabine, T., Straf, M., Tanur, J., and Tourangeau, R. (1984), *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*, Washington, DC: National Academy Press.

Lessler, J., Salter, W., and Tourangeau, R. (1989), "Questionnaire Design in the Cognitive Research Laboratory: Results of an Experimental Prototype," *Vital and Health Statistics, Series 6, No. 1* (DHHS Publication No. PHS 89-1076), Washington, DC: U.S. Government Printing Office.

Martin, E. (1987), "Some Conceptual Problems in the Current Population Survey," in *Proceedings of the Section on Survey Methods Research*, American Statistical Association, pp. 420-424.

McKay, R. (1990), "Application of Persuasive Communication Strategies to a Business Establishment Survey," in *Proceedings of the Section on Survey Methods Research*, American Statistical Association, to appear.

Miller, L. A. and Downes-LeGuin, T. (1989), "Reducing Response Error in Consumers' Reports of General Expenses: Application of Cognitive Theory to the Consumer Expenditure Interview Survey," *Advances in Consumer Research*, in press.

Mullin, P. (1990), "Proposal for Laboratory Research on the Feasibility of an Extended Interview Period for the CPS," unpublished memorandum to A. Tupek, in preparation.

Norwood, J. and Dipbo, C. (in press), "Government Applications," in *Questions about Questions: Memory, Meaning, and Social Interaction in Surveys*, New York: Russell Sage.

Phipps, P. (1990), "Applying Cognitive Techniques to an Establishment Mail Survey," paper to be presented at the annual meeting of the American Statistical Association, Anaheim, California, August.

Royce, J.R. (1973), "The Present Situation in Theoretical Psychology," in B.B. Wolman (ed.), *Handbook of General Psychology*, Englewood Cliffs, NJ: Prentice Hall.

Shoemaker, H., Bushery, J., and Cahoon, L. (1989, in press), "Evaluation of the Use of CATI in the Current Population Survey," in *Proceedings of the Section on Survey Research Methods, American Statistical Association*.

Squire, L. (1987), *Memory and Brain*, New York: Oxford University Press.

Suchman, L. and Jordan, B. (1990), "Interactional Troubles in Face-to-Face Survey Interviews," *Journal of the American Statistical Association*, 85, 232-240.

Tanur, J. (1990, in press), "Reporting Job Search Among Youths: Preliminary Evidence from Reinterviews," in *Proceedings of the 1990 Annual Research Conference*, Washington, DC: U.S. Department of Commerce, Bureau of the Census.

Tourangeau, R. (1984), "Cognitive Sciences and Survey Methods," in *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*, T. Jabine, M. Straf, J. Tanur, and R. Tourangeau (eds.), Washington, DC: National Academy Press.

Tucker, C. and Bennett, C. (1988), "Procedural Effects in the Collection of Consumer Expenditure Information: The Diary Operations Test," in *Proceedings of the Section on Survey Methods Research, American Statistical Association*, pp. 256-261.

Tucker, C., Miller, L., Vitrano, F., and Doddy, J. (1989), "Cognitive Issues and Research on the Consumer Expenditure Diary Survey," paper presented at the annual American Association for Public Opinion Research Conference.

Wolkowitz, O.M. and Weingartner, H. (1988), "Defining Cognitive Changes in Depression and Anxiety: A Psychobiological Analysis," *Psychiatry Psychobiology*, 3, 1-8.

Wright, P. (1980), "Strategy and Tactics in the Design of Forms,"
Visible Language, XIV 2, pp. 151-193.

THE ROLE OF A COGNITIVE LABORATORY IN A STATISTICAL AGENCY

Monroe G. Sirken
National Center for Health Statistics

Introduction

The statistical survey is an invention of the twentieth century. It produces a commodity, namely information, which many believe is the most important property in the modern world. Our Federal establishment, for example, would be unable to function nearly as effectively without the information being produced by surveys that are conducted by the Federal agencies represented at this Seminar. The Congressional and Executive branches use Federal surveys to monitor the nation's well-being, to evaluate the government's social, health and economic programs, and to plan legislation involving the collection of billions of tax dollars and the disbursement of billions of benefit dollars. Federal surveys could not have attained this level of acceptance and importance without the technological advances in survey methods that have occurred during the past half century. However, we can hardly afford to be complacent. As data producers, we are even more mindful than data consumers of the limitations of current survey technology. We realize that further technological advances are essential to assure that Federal surveys will meet the growing needs for more and better survey data.

There have been two major technological advances in survey methodology during the past 50 years and I believe a third may be in the offing. Each advance has introduced innovative technologies for improving the precision of the survey measurement process and was made possible by technology and theory transfers from the applied sciences. The "sampling" revolution in survey methodology that began in earnest during the 1930's came about as a result of technology transfers from the statistical sciences, and produced substantial advances in survey sampling and estimation methods. The "automation" revolution had its onset in the late 1960's. It came about as a result of technology transfers from the computer sciences, and has produced substantial advances in the methods of compiling and processing survey data. The "cognitive" revolution, which, as some of us believe got underway during the 1980's [Jabine, 1989], was made possible by technology and concept transfers from the cognitive sciences. Whether called a revolution or a movement, it has been introducing improved methods of designing data collection instruments and conducting questionnaire design research.

Federal Statistical agencies were major players in the "sampling" and "automation" revolutions in survey technology. Now they are playing a major role in the "cognitive" movement by

developing and applying cognitive laboratory techniques to find better solutions to survey response problems. It is noteworthy that the cognitive movement is not confined to the U.S. government nor to the United States [Jobe and Mingay, 1991]. This paper, moreover, deals with only one part of the U.S. movement, namely, the work of the cognitive laboratory at the National Center for Health Statistics. The paper briefly describes the history and programs of the NCHS Laboratory and outlines the Laboratory's benefits to survey research, cognitive psychology, and Federal statistics.

History of the NCHS Laboratory

Until 1984, the role of cognition in the survey measurement process was largely ignored in the survey research programs of the National Center for Health Statistics. None of the earlier NCHS projects had been conducted in a cognitive laboratory, though one study [Laurent, Cannel and Marquis, 1972] used psychological theories to guide the development of interviewer and questionnaire techniques. Prior to 1984, survey response had been modeled as a two stage stimulus/response process with little attention paid to the effects that the respondents' mental processes had on the accuracy of their responses. In accordance with this psychological paradigm, survey research investigated the error effects of survey instruments and procedures almost exclusively in field tests. Since these field tests sought to replicate the actual conditions of the survey, they provided little opportunity to investigate cognitive issues, such as the following:

- What kinds of cognitive processing modes and strategies do respondents use in answering survey questions?
- How do the cognitive processing modes and strategies of survey respondents affect the accuracy of their responses to survey questions?

In 1984, with the support of an NSF grant, the NCHS embarked on a demonstration project that was motivated largely by the work of the Advanced Research Seminar on the Cognitive Aspects of Survey Methodology [Jabine, Straf and Tanur, 1984]. This project sought to demonstrate the utility of investigating the cognitive aspects of answering survey questions in a laboratory setting as a means of improving the design of Federal survey instruments [Sirken and Fuchsberg, 1984]. The project compared alternate versions of the dental supplement to the questionnaire of the 1986 National Health Interview Survey. One supplement was designed by the traditional field test method and the other by the proposed cognitive laboratory method [Lessler and Sirken, 1985].

The rationale for the demonstration project as expressed in the NSF grant proposal [Sirken, 1984] was:

"...because (1) questionnaire design is one of the weakest links in the survey measurement process, (2) past efforts to improve the quality of questionnaire have posed serious and difficult methodological problems, (3) the traditional field methods currently being used to improve questionnaire design are inadequate by themselves to handle many of these problems, and (4) complimentary methodologies that are not subject to the weakness of traditional field methods need to be developed, it is [therefore] essential to investigate the potential of using the [combined] techniques of the statistical and cognitive sciences in a laboratory setting as a complementary methodology for improving questionnaire design..."

The demonstration project was conducted in an interdisciplinary mode and in close collaboration with university scientists so that, as the NSF grant proposal noted, another potential benefit was:

"...it could go a long way in bridging the gap that exists between cognitive scientists academia and survey statisticians in Federal Statistical Agencies..."

This was critical to the ultimate success of the project because it was felt that gap between the disciplines had been largely responsible for the delay in applying cognitive methods in survey research.

At the successful conclusion of the demonstration project in 1986, NCHS established, with the support of a second NSF grant, the National Laboratory for Collaborative Research in Cognition and Survey Measurement. The National Laboratory's broad mission is to promote and advance interdisciplinary research on the cognitive aspects of survey methodology among Federal Statistical Agencies and the nation's universities and research centers. Interdisciplinary research with university scientists is promoted by a Collaborative Research Program which awards competitive research contracts and appoints visiting scientists. Collaborative research with other Federal Agencies is promoted by the Questionnaire Design Research Laboratory which serves as the workplace for NCHS and other Federal Agencies to conduct intramural research [Royston, et al 1986]. The Collaborative Research Program has been largely funded by NSF grants and the Questionnaire Design Research Laboratory has been partially funded by reimbursable work agreements with other PHS Agencies [Sirken, et al 1990].

Activities of the NCHS Laboratory

Much of the work of the National Laboratory is based on a cognitive theory of survey response errors that can be stated as

follows: "survey respondents carry-out a series of mental tasks in the interval between being asked a survey question and providing a response. When these mental tasks pose serious mental burdens for respondents they are likely to cause response errors." This view of the survey response process stimulated the development of cognitive methods for designing and pretesting questionnaire and for conducting questionnaire design research. Developing and testing survey instruments has short term objectives, namely, to detect and revise the design flaws before the survey instruments are field tested. In contrast, questionnaire design research objectives are long term, namely, to improve the designs of the next generation of survey instruments. These differences in objectives led to the development of distinctly different cognitive methods for developing and testing survey instruments and for conducting questionnaire design research.

Developing and Pretesting Questionnaires

The cognitive laboratory approach to developing and pretesting survey questionnaires is based on the premise that difficult, unreasonable or impossible the mental tasks implicit in some survey questions increase the likelihood of response errors. For example, survey questions containing terms respondents do not understand, that are vague or ambiguous, that impose unrealistic demands on recall, that require complicated mental calculations, that contain too many elements for the respondent to think about simultaneously, that involve issues the respondent knows or cares little about, or that ask for embarrassing or threatening information-all impose cognitive burdens that are likely to result in invalid responses.

The realization that questionnaires obtain poor quality data when they ask respondents to perform difficult, if not impossible, mental tasks led to the development of a battery of laboratory techniques for investigating the cognitive burdens posed by survey questions [Bercini, In press, Royston, 1989] including think-aloud interviews, in-depth probing and focus group discussions, etc. These techniques are not new to questionnaire designers [DeMaio, 1983] but never before had they explicitly and systematically served as means of observing the manner in which respondents mentally process survey questionnaires and procedures.

Intensive interviewing techniques detect questionnaire design flaws by observing the cognitive problems that result from these flaws. Poor questionnaire designs may impose difficult mental tasks at any cognitive stage of the response process including comprehending the questions recalling or estimating the information needed to answer the questions, and deciding whether or how to answer the questions. Identifying the underlying cognitive difficulties experienced by respondents facilitates the process of revising the questionnaires appropriately.

Many questionnaire design problems detected and repaired by laboratory techniques are far less likely to be detected by traditional field testing methods. Consider the following question which was proposed for the National Health Interview Survey (NHIS), "During the past 12 months, have you been bothered by pain in your abdomen?" When laboratory respondents were asked this question, most answered it readily with a "Yes" or a "No." It was not until the laboratory interviewer probed into how respondents interpreted the term "abdomen" that it became apparent that respondents were unsure of what section of the body to include. The interviews also determined that respondents had variable interpretations of the phrase, "bothered by," which in turn, affected whether they answered the question affirmatively or negatively. Intensive interviewing methods not only revealed that the question was apt to result in response errors, but also the underlying cause of the problem. When the cause of a question problem is understood, the solution is more likely to be found. In this case, part of the solution was a respondent flash card that showed an outline of the torso with the abdominal area shaded in.

Intensive interviews are conducted by laboratory trained questionnaire designers with many years of survey research experience. Paid subjects are recruited for the interviews. The topic and target populations of the survey determine the criteria for subject recruitment. Subjects are often selectively recruited to include those that would be most burdened by the survey questions or least successful in adopting effective mental strategies in answering the questions. Laboratory testing is usually carried out in interviewing waves of 5 to 10 subjects at a time; the questionnaire is revised in consultation with the sponsor after each wave; and the testing continues until an acceptable version is obtained. Typically, flawed questions undergo 2-4 revisions before an acceptable version is ready for field testing. Field testing is essential in order to determine how the questionnaire will work under actual survey conditions. Additional laboratory testing may be needed to evaluate the questionnaire revisions that are suggested by the field test.

Depending on the complexity and scope of the questionnaire and on the number of conceptual problems associated with it, laboratory testing can be completed within several weeks or could span a longer period. For example, projects that involve special subject recruitment and testing may require a lead time of about six months or even longer. Also, laboratory projects are conducted collaboratively with survey sponsors and therefore involve frequent meetings to assure that the designed questionnaires satisfy the sponsors' research objectives.

Questionnaire Design Research

Cognitive methods of conducting questionnaire design research investigate why some survey questions and procedures pose cognitive tasks that are difficult, unreasonable or impossible for respondents to perform. In the same way that much has been learned in medicine by studying the cognitive aspects of amnesia and other memory disorders, so it is hoped that much can be learned in survey research by studying the cognitive aspects of questionnaires that pose severe response burdens.

Questionnaire research seeks to improve the design of the next generation of survey questionnaires, especially those questionnaires dealing with topics for which better quality survey data are needed. Causal relationships between the mental tasks performed by respondents and the accuracy of their responses are investigated in experiments. These experiments may be conducted in the cognitive laboratory or embedded in on-going surveys. The laboratory approach makes it possible to undertake many types of complex experiments that would be administratively impossible or prohibitively expensive to conduct as field experiments. Embedding cognitive experiments in on-going surveys makes it feasible to test laboratory findings under actual survey conditions.

Several features of cognitive laboratory experiments are noteworthy. They are interdisciplinary, involving the joint participation of cognitive psychologists and survey researchers. They generally involve testing questions that ask for the kinds of information that typically is poorly reported in surveys. They investigate those mental tasks implied by the survey questions that pose the greatest risks to accurate reporting. For example, if the question implied retrospective reporting, the focus would be on the cognitive aspects of the memory tasks and if the question asked for sensitive information the focus would be on the cognitive aspects of risk taking under conditions of uncertainty.

Generally, the subjects of laboratory experiments are recruited from population frames that contain information needed to validate the experiment's findings. For example, the laboratory subjects for experiments on retrospective reporting of medical visits were selected from the files of a Health Maintenance Organization. Because the files provided access to the recruitment of subjects with known health conditions and doctor visit patterns [Means, et al, 1988]. Finally, the findings of the laboratory experiments are interpreted in terms of their potential contributions to cognitive theory as well as their implications for improving the design of survey instruments.

A recent project on dietary recall in nutrition surveys illustrates some of the benefits of conducting experiments in a cognitive laboratory. This complex multi-experiment project, involving randomization of subjects, diary keeping, and multiple

data collection sessions, could probably not have been undertaken as a traditional field experiment. The project investigated the cognitive burdens posed by the kinds of questions that are asked in household nutrition surveys [Smith, In press]. Generally these surveys collect dietary histories, food frequency inventories, and data on food portion sizes. Collecting these kinds of data imposes mental tasks involving free recall, frequency estimation, and magnitude estimation, respectively. Separate laboratory experiments were designed and conducted to assess the ability of respondents to provide accurate information on each of these tasks. The laboratory subjects participating in these experiments kept food diaries so their subsequent responses to dietary questionnaires could be validated.

For example, one of the nutrition survey experiments tested the effect of varying the portion size definitions on respondents' reports of the amount of food consumed. For each listed food item, respondents indicated whether their typical portion was small, medium or large in comparison with a defined medium portion size. Surprisingly, the food consumption reports in the experiment were invariant to changes in the definition of medium portion size. These findings raise serious questions about the design of nutrition survey questionnaires and the quality of survey data on food consumption that are based on portion size reports.

Over the past several years, laboratory experiments have investigated the cognitive factors involved in responding to difficult-to-answer questions on a variety of health related topics including utilization of health services, cigarette smoking histories, illegal drug use, chronic pain episodes, and chronic disease prevalence.

A recent project on recall of doctor visit illustrates the benefits of embedding experiments in surveys. This split-ballot experiment was embedded in the pilot study of the National Medical Expenditure Survey. The experiment investigated the relative accuracy of retrospectively reporting doctor visits in a forward or in a backward temporal order [Jobe, et al, 1990]. It was suggested by the findings of previous laboratory experiments indicating that subjects varied in their preference between forward and backward recall order but that backward recall seemed to produce more accurate reporting [Loftus, 1985].

The survey experiment assessed the accuracy of forward, backward and free recall reporting strategies by comparing the medical visits reported by each strategy with the visits listed in medical records. The survey experiment did not confirm the findings of the laboratory experiments and showed little difference in accuracy between the alternative recall strategies. It was concluded that there was no evidence to suggest that survey instruments should be designed to favor either the forward, backward or free recall strategies.

Cognitive experiments involving survey material, whether conducted in laboratories or embedded in surveys, are valuable for several reasons. First, they provide in-depth knowledge about the cognitive processes respondents use in answering hard-to-answer survey questions. In particular, they often identify the kinds of question approaches that pose response burdens. And they suggest methods of designing the questionnaires to reduce the response burdens and response errors. Secondly, because validation information is almost always collected (e.g., diaries, medical record matches, and biochemical markers) the response error effects of different questionnaire designs and cognitive strategies can be assessed. Third, the cognitive bounds on the abilities of respondents to perform specified kinds of mental tasks (comprehension, recall, etc.) posed by survey questions can be assessed.

Benefits of the NCHS Laboratory

The activities and programs of the NCHS cognitive laboratory during the past five years have benefitted survey research, cognitive science and Federal statistics in variety of ways. Some of the benefits are briefly outlined in these summary remarks.

Survey research has benefitted from the development of methods for investigating the cognitive aspects of the survey response process. Intensive interviewing methods were perfected for designing and pretesting survey instruments in a laboratory setting, and experimental methods were perfected for conducting laboratory experiments and for embedding experiments in on-going surveys.

Cognitive science benefitted from the opportunities afforded its scientists by the NCHS laboratory to participate in the interdisciplinary research projects in cognition and survey measurement. Cognitive psychologists participating in these projects had opportunities to test cognitive theories with real world survey phenomena either in laboratory experiments or in experiments embedded in on-going surveys. And it is believed that the gains in cognitive psychology will ultimately benefit survey research and the quality of Federal surveys.

The activities of the NCHS laboratory fostered an appreciation and respect for the importance of conducting cognition and survey measurement research within and outside the Federal establishment. For example, the NCHS laboratory played a vital role in designing and testing NCHS survey instruments during the past several years, and it is being viewed increasingly as a PHS laboratory with a mission to service the needs of agencies throughout the Public Health Service. As the first cognitive laboratory of its kind devoted to survey research, the NCHS laboratory served as a point of reference, if not the prototype, for the cognitive laboratories

that have since been established at other statistical agencies including the Bureau of the Census, Bureau of Labor Statistics and Statistics Sweden. Information dissemination has always been a high priority activity and during the past five years, the NCHS laboratory staff and collaborators published nearly 50 reports, and presented more than 100 papers at meetings and conferences.

Whether the existing movement in cognition and survey research, of which the NCHS laboratory is a part, will evolve into a full-fledged cognitive revolution with an impact equal to the sampling and automation revolutions remains to be determined. We will know that the cognitive revolution has occurred when it becomes apparent that the cognitive sciences are providing scientific support to survey response research comparable to the support the statistical and computing sciences have been providing to research in survey sampling and in the automation of survey data.

References

Bercini, D.H. Presented at the EPA/AOWNA Symposium on Total Exposure Assessment Methodology. Pretesting Questionnaire in the Laboratory: An Alternative Approach. In print. Toxicology and Industrial Health.

DeMaio, Theresa J. (Ed.) (1983). Approaches to Developing Questionnaires. Statistical Policy Working Paper 10. Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget. Washington, D.C.

Jabine, Thomas B. (1990). Cognitive Aspects of Questionnaire Development. Presented at the EPA/AOWNA Symposium on Total Exposure Assessment Methodology. In print. Toxicology and Industrial Health.

Jabine, T.B., Straf, M.L., (1984). Tanur, J.M. and Tourangeau, R. (Eds.). (1984). Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines. Washington, D.C. National Academy Press.

Jobe, J.B., White, A.A., Kelley, C.L., Mingay, D.J., Sanchez, M.J., and Loftus, E.F. (1990). Recall Strategies and Memory for Health Care Visits. Milbank Memorial Fund Quarterly/Health and Society, 68, 171-189.

Laurent, A.C., Cannell, C. and Marquis, K. (1972). Reporting Health Events in Household Interviews: Effects of an Extensive Questionnaire and Diary Procedure. Vital and Health Statistics, Series 2, No. 49 (DHHS Publication No. PHS 91-1079). Washington, D.C., U.S. Government Printing Office.

Lessler, J.T. and Sirken, M.G. (1985). Laboratory-Based Research on the Cognitive Aspects of Survey Methodology: The Goal of the National Center for Health Statistics Study. Milbank Memorial Fund Quarterly/Health and Society, 63, 565-581.

Loftus, E.F., and Fathi, D.C. (1985). Retrieving Multiple Autobiographical Memories. Social Cognition. Vol. 3, pp. 280-95.

Royston, P.N. (1989). Using Intensive Interviews to Evaluate Questions. In F.J. Fowler, Jr. (Ed.), Heath Survey Research Methods, (pp. 3-7). (DHHS Publication No. PHS 89-3447). Washington, D.C., U.S. Government Printing Office.

Royston, P.N., Bercini, D.H., Sirken, M.G. and Mingay, D. (1986). Questionnaire Design Research Laboratory. American Statistical Association, 1986 Proceedings of the Section on Survey Methods Research, pp. 703-707.

Sirken, Monroe G. (1986). National Laboratory for Collaborative Research on Cognition and Survey Measurement. Grant Proposal to the National Science Foundation. Washington, D.C.

Sirken, Monroe G. (1984). Laboratory-Based Research on the Cognitive Aspects of Survey Methodology. Grant Proposal to the National Science Foundation. Washington, D.C.

Sirken, M.G. and Fuchsberg R. (1984). Laboratory Based Research on the Cognitive Aspects of Survey Methodology. In Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines. Washington, D.C. National Academy Press.

Smith, A.F. (in press). Cognitive Processes in Long-term Dietary Recall. Vital and Health Statistics, Series 6, No. 4 (DHHS Publication No. PHS 91- 1079). Washington, D.C., U.S. Government Printing Office.

DISCUSSION

Elizabeth Martin
U.S. Bureau of the Census

In their two papers, Monroe Sirken of the National Center for Health Statistics, and Cathryn Dippo and Douglas Herrmann of the Bureau of Labor Statistics, document the activities of the cognitive laboratories which were established in 1984 and 1988, respectively, at their two agencies. The cognitive laboratories represent a commitment to survey data quality which is a credit to the two agencies. And Monroe Sirken and Cathryn Dippo, as two of the main instigators and initiators responsible for establishing the laboratories, deserve credit and appreciation for their effort and accomplishment. The record of achievement by the two laboratories is a good one. Dippo and Herrmann organize their paper around a clear and comprehensive discussion of the sources of cognitive problems which can introduce errors in the response process; it is impressive how many of these problems have already been tackled in the BLS Collection Procedures Research Laboratory in its short history. Excellent research on a range of topics is also being conducted at the NCHS National Laboratory for Collaborative Research in Cognition and Survey Measurement, though in his paper Sirken does not actually describe the research. The NCHS lab lives up to the "collaborative" in its name; the number and caliber of academic researchers who have been involved in their projects are very high.

The growth of laboratory-based research on cognitive aspects of survey methodology is described by Dippo and Herrmann as a "movement" and by Sirken as a "revolution." These characterizations accurately reflect the enthusiasm and ferment of activity and new ideas in this area. However, "revolution" may not be the most useful metaphor to describe how cognitive psychology is affecting (or, more importantly, should affect) survey research. In fact, the metaphor of "revolution" reflects and reinforces a weakness of the work currently going on in the new cognitive laboratories.

By emphasizing discontinuity with the past, researchers are led to ignore relevant work which preceded many of the methods and ideas of the current "movement." Sirken characterizes survey research as (until recently) "based almost exclusively on the behaviorist paradigm" with "respondent's mental states ... virtually ignored." This isn't accurate. Survey researchers, at least those practicing in academic or commercial settings, have hypothesized about and investigated psychological states intervening between survey questions and respondents' answers at least since World War II. (Jean Converse's Survey Research in the United States: Roots and Emergence, 1890-1960 provides a fascinating and useful history which traces the intellectual

origins of survey research.) Much of this work is still very relevant, and should be built on rather than ignored. For example, Dippo and Herrmann state that, "except for social desirability, the survey field is just beginning to investigate factors that affected communication of responses." They would benefit from reviewing the survey literature on the topic of communication, beginning with Herbert Hyman et al's comprehensive, Interviewing in Social Research, published in 1954. The methods used in the cognitive laboratories also have roots in the past. For example, Naomi D. Rothwell used very similar methods to conduct research on questionnaire design at the Census Bureau during the 1960s and 1970s. It is a bit of an overstatement for Sirken to claim in his paper to have invented the cognitive laboratory, without acknowledging similar, earlier activities.

In the field of survey research, there is a tradition of applying ideas from psychology to survey measurement issues. For the new work in the cognitive laboratories to advance the state of the art of survey measurement, it should build on this tradition. This would also increase its credibility to many survey researchers.

A danger of the "revolution" metaphor is it suggests a philosophy of "out with the old, in with the new." In some cases, this leads researchers to forget what they know about good survey practice. Compared to a survey, the cognitive laboratories generally rely on more intensive, less structured interviews with smaller numbers of respondents. This approach can be very informative about the nature and sources of cognitive errors in surveys. However, the "samples" usually are very small and not selected according to probability methods. One must be cautious drawing inferences from the results of most of the cognitive lab studies to date. For instance, I think Dippo and Herrmann are overstating the case when they conclude that, "research done at BLS shows clearly that proxy recall is different than self recall, both in terms of amount and kinds of information recalled." Laboratory findings such as this are more usefully thought of as hypotheses which should be subjected to more rigorous testing in a sample survey, and/or experimentally.

It is important to keep in mind that standards of evidence and proof still apply to research conducted in the cognitive laboratories. In some writings, the word "cognitive" is repeated so often as to suggest that the writer believes the word itself is sufficient to establish the merits of the research. But the researcher is still obliged to make his or her case on the evidence. For example, Sirken presents an example of a question on marijuana use which he says was improved by cognitive testing. How do we know it is better? He presents no evidence or logic to support his claim. In the long run, if the cognitive "movement" is to be taken seriously, it must demonstrate, not simply assert, the

value of its products, and be wary of the temptation to oversell itself.

I believe there are two common goals behind the activities in the cognitive laboratories. One goal is to improve particular survey measurements. The second is to develop a theoretical foundation (beyond sampling theory) for improved survey design. The latter, broader aim requires that we develop better measures of nonsampling errors, and a better understanding of the effect of alternative survey designs on nonsampling errors. Methods and ideas from cognitive psychology are tools for achieving both specific and general improvements, but are not an end in themselves. Other social sciences (for example, social psychology) also have relevant knowledge to contribute.

With these goals (and the previously-stated cautions) in mind, what then is new and revolutionary about the work being done in the cognitive laboratories? First, this research has yielded new appreciation of the vulnerability of factual survey questions to biases and errors. I think it is fair to say that most government statisticians and academic survey methodologists probably have taken for granted the validity of simple factual questions. The research on problems of comprehension, recall, and other cognitive difficulties is contributing to a more sophisticated understanding of how much we have yet to learn about the error properties of survey measurements. Second, and more important, the research in the cognitive labs represents a new and more extensive set of methods for pretesting survey questionnaires and procedures. This in itself is a great leap forward. Traditionally, pretests of survey questionnaires have been ad hoc and informal, based on interviews with a few respondents and with no real guidelines beyond common sense to decide when one has succeeded or failed. The cognitive laboratories are changing that. Close and in-depth examination of problems of respondent comprehension, recall, and judgment, is shedding new light on the causes of these problems and (better yet) new ideas about how to correct them. The new methods which are being used and developed in the cognitive laboratories form a logical series of pretests prior to fielding a survey, proceeding from intensive, informal interviews, to small-scale experiments testing alternative questions or designs, to large-scale field experiments. In addition, as Cathryn Dipbo points out in her remarks, testing can be integrated into the main survey itself, to provide ongoing information about nonsampling errors. The new methods thus make possible a more scientific and systematic approach to pretesting, and they promise to yield improvements in the quality of data collected by the federal government.

DISCUSSION

Murray Aborn
National Science Foundation (retired)

I am grateful to my co-discussant, Elizabeth Martin of the Census Bureau, for providing the perfect lead-in to my own commentary on the papers presented at this session. Dr. Martin reminded us of the importance of viewing any disciplinary development from the perspective of its historical predecessors, and in this connection she succeeded in moving the advent of CASM (Cognitive Aspects of Survey Methodology) -- writ large -- back several decades from the year most commonly cited as the date of its birth -- namely, 1980.

More consequential than revising our perception of the chronology of CASM (again writ large) is the difference Dr. Martin's remarks point up between the characterization of CASM in the paper presented by Cathryn Dipbo and Douglass Herrmann of the Bureau of Labor Statistics, and the one presented by Monroe Sirken of the National Center for Health Statistics. Dr. Martin's remarks implicitly characterize CASM as a reawakening of old concerns, and thus place her in strong agreement with Dipbo's and Herrmann's labeling of CASM as a "movement," in contrast with Sirken's labeling of CASM as a methodological "revolution." Indeed, there is much to support the view of CASM as a movement; for instance, the enthusiasm of its adherents and the growing frequency with which its ideology is being endorsed by sectors of the statistical community and users of statistical data generally who have heretofore tended to ignore the psychosocial underpinnings of survey-taking (see, for example, Suchman and Jordan, 1990).

However, this does not mean that Sirken's description of CASM as representing a revolutionary development is totally incorrect. It may merely be premature, for the potential of CASM as a true breakthrough -- as a true revolution in survey research -- is clearly present in the programmatic and research agenda laid out for it in the seminal CASM document prepared by the National Academy of Sciences (see Jabine, et al, 1984). At the present time, only half the CASM prospectus is being actively pursued; namely, those objectives having to do with the adoption of certain recent advances in cognitive science into the survey design and instrumentation process. What we have seen little of to date is action on those objectives having to do with the use of surveys as naturalistic test beds for laboratory-based theories of the functioning of the neuronal mind and, ultimately, the emergence of a new paradigm for social/behavioral research in which survey-taking plays an important role in understanding such basic cognitive phenomena as how the brain stores memories and how mental imagery influences perception and recall, and in which developments in cognitive science relating to such branches of the field as

natural language semantics are used to produce greatly improved methods for achieving high-quality survey measurement. In other words, fulfillment of the "cognitive revolution" alluded to in Monroe Sirken's paper is clearly in prospect, but is yet to materialize.

I shall have a bit more to say on this subject at the close of my commentary; meanwhile, however, it is my opinion that much of the force behind Dr. Martin's view of CASM as a reawakening of old survey concerns -- as a "movement" more so than a "revolution" -- stems from the present truncated status of the programmatic agenda initially prescribed for the field. This gives CASM the appearance of a one-sided effort to adopt, in fairly superficial terms, some of the investigative techniques employed in recent laboratory-based cognitive psychology, and incorporate them in the conventional procedures for constructing and pretesting survey questionnaires. Under such a perspective, not much may appear to have been added to what has long been known to be of influence in survey responding, and audiences such as the one attending the present session may rightfully feel that CASM amounts to little more than another real-life example of the familiar tale of "The Emperor's New Clothes" which, albeit a story from the literature of childhood, embodies a profound adult theme concerning human gullibility and our tendency to accept uncritically what experts -- genuine and otherwise -- tell us is true, novel, or significant.

Now, let me examine the Emperor's New Clothes proposition against the CASM-engendered activities at the BLS and NCHS laboratories reported in the papers by Dipppo and Herrmann and by Sirken. Reducing a sample of these activities to their most generic properties (in the sense of survey factors which induce

response error), I would break them down into the following classification:

COLLECTION PROCEDURES
RESEARCH LABORATORY
(BLS)

- Question Ambiguity
(The extent to which a question may be interpreted in more than one way.)
- Long-term Recall
(The length of time over which the respondent is required to retrieve from memory.)
- Emotional Loading
(The degree of psychological stress which a question may place upon the respondent.)
- Subcultural Norms
(Question comprehensibility across ethnic subgroups.)
- Social Desirability
(The extent to which a question is likely to elicit a normative rather than an idiopathic response.)

QUESTIONNAIRE DESIGN
LABORATORY
(NCHS)

- Question Wording and Order
(The differential results induced by synonymous variation rearrangement of sequence.)
- Memorial Decay
(The validity -- or veridicality -- of information supplied from short- and long-term memory.)
- Affective Sensitivity
(The likelihood that a question may be embarrassing or impinge upon the respondent's privacy.)
- Linguistic Complexity
(The effect of grammatical construction on the respondent's ability to comprehend.)
- Lexical Level
(The extent to which a question requires the respondent to have specialised -- in this case medical -- knowledge.)

Now, it is hard to believe that the many survey researchers trained in social psychology and cognate fields of social science are oblivious to influences -- such as those charted above -- regardless of whether intellectual, technical, and/or cost factors make it impractical to subject such nonsampling sources of error to adequate control, or to estimate the proportion of total survey error due to their ubiquitous presence.

To take the phenomenon of Social Desirability, for example, it does not require a social scientist to comprehend the universal tendency of people to present a societally acceptable facade when questioned about attitudes and behavior. The popular press and many humorous books have for decades poked fun at surveys by ridiculing the informational value of asking such survey items as, "Do you bathe at least once a week?" or "Do you brush your teeth every day?"

To take some other examples, did it require CASM to alert survey researchers to the difference in results when a question is phrased one way as opposed to another? Or to the difficulty of most respondents to deal with questions presented in grammatically complex form? Or to the impingement of certain areas of questioning on the sensitivity of respondents? Or to memorial decay over time? Or to a respondent's understanding of questions embodying medical terminology?

I can't resist regaling the audience with a personal anecdote illustrating how ordinary, and even old-fashioned, if you will, is appreciation of the fact that few individuals not trained or highly educated in medicine cannot comprehend medical lexicography, and that one is apt to get ludicrous results from asking questions embodying medical terminology.

More than 25 years ago, when employed at the National Institute of General Medical Sciences, I shared an office with a public health epidemiologist who had just returned from a tour of duty in Puerto Rico. He told me of an effort to obtain data on the extent of interruption to normal life activities due to amoebic dysentery, which was then prevalent in most rural areas of Puerto Rico. Having never before conducted a survey, his group of public health officials put together a series of questions utilizing such terms as diarrhea and defecation to get estimates of frequency. When the obtained results showed an average of only one to two bowel movements per day, the survey takers knew something was wrong and quickly realized that it was likely due to the language employed in identifying the disease.

The Public Health people reran a small subsample of respondents using the term "bowel movement" in the questionnaire, and obtained a slightly higher, but still medically incredible, estimate of frequency. Finally a native informant suggested that they phrase all questions pertaining to diarrhea in terms of La Mange -- or "The Curse" as it was known in the rural areas of the island -- and when they did this, the average reported frequency shot up to a more medically believable 11 or 12 occurrences per day.

If sheer knowledge of the fact that such variables as level of lexical comprehension, differences in subcultural norms, and the tendency to respond in socially desirable ways are sources of error in survey research, what, then, is it that is truly new about the CASM movement?

There are, to my mind, three major issues that have been brought to the fore by the CASM movement, coupled to the addition of new technical procedures which have proved powerful in cognitive research in psychology and artificial intelligence. And, as I have mentioned before and will emphasize at the close of my remarks, there is the potential for bringing about a truly interdisciplinary

effort to understand just what goes on in the interactional dynamics for survey and respondent.

The three major issues which have surfaced as a result of CASM are:

1. A reawakening of the essential conflict between survey questionnairing and ordinary conversation owing to the need for artificially imposed standardized conditions of administration from the standpoint of survey statistics on the one hand, and the natural world existence of individual differences in mentality on the other.
2. The extent to which laboratory-based treatments and results can be transferred to the field in the case of survey-taking. This issue is of general importance to social science, as well as being particularly relevant to survey research insofar as the laboratory setting, which provides greater conditions of control and flexibility, creates possibilities for a more systematic approach to instrumentation, and hence to survey measurement.
3. The degree to which the contemporary shift in the underlying paradigm of survey research's cognate substantive discipline -- i.e., psychology -- requires a realignment away from behaviorism and toward cognition. CASM represents a bold attempt to test this issue and assay its yield, but there has thus far been far too little involvement of cognitive psychology per se apart from the importation of certain investigative techniques.

I by no means wish to detract from the accomplishments reported in the papers by Dipbo and Herrmann and by Sirken based upon the importation of the techniques employed in contemporary cognitive psychology, into the innovative laboratory facilities now ensconced in such two prestigious governmental agencies as BLS and NCHS. Much thought and expertise have been applied to the transfer of technology represented by the successful adoption of such cognitive probes and methods as: (1) Focus Groups; (2) Part-set Cueing; (3) Protocol Analysis; and (4) Think-aloud procedure.

But in my opinion, this could be just the beginning of a truly revolutionary development in survey research and, through its influence, on social science more broadly. The laboratory-based techniques and procedures you have heard presented at this session are derived from research begun in the early 1960's by Nobel Laureate Herbert Simon and Alan Newell that resulted in the General Problem Solver and led to the foundations of the field of artificial intelligence (Barr and Feigenbaum, 1982). The more recent work of Simon (Simon, 1987), shows the even greater potential of cognitive technology to uncover human information processing systems.

However, there is reason to be both pessimistic and optimistic about the future of CASM. On the one hand, the statistical framework of survey research -- the dominant framework for the field -- is concerned with drawing inferences about populations -- about whether the sample of a population is large and representative enough to permit accurate and valid conclusions to be reached about the distribution of characteristics in the population from which the survey sample was drawn. On the other hand, the cognitive framework is concerned with drawing accurate and valid inferences about individuals -- about respondent "truthfulness," if you will.

Therefore, one framework calls for instrumentation designed to enhance person-to-person comparability, while the other calls for instrumentation designed to enhance the assessment of person-to-person variations on each survey variable.

It is the work of the two survey/cognitive research laboratories reporting here today that represents one of the two reasons I find to be optimistic about the future of CASM. Such facilities offer the best opportunities for reconciling the conflicting survey conceptual frameworks described above.

The other reason I find to be optimistic lies in the pronouncement appearing in a neuropsychological book which has become a national bestseller in addition to its importance to the scientific literature on brain-behavior relationships. I refer to -- and endorse to you as top-quality literature as well as a work of cognitive science importance -- Oliver Sacks' The Man Who Mistook His Wife for a Hat. I close my remarks by quoting from a passage in this work that, I believe, should stimulate cognitive scientists to become fuller participants in CASM, recognizing that survey centers and facilities are ideally suited to cognitive explorations and offer the prospect of a vital new interdiscipline.

After presenting and analyzing the case of The Man Who Mistook His Wife for a Hat, Sacks concludes, as I do here, that:

Our cognitive sciences are themselves suffering from an agnosi similar to the one afflicting the man who mistook his wife for a hat. That man may thus serve as a warning and parable of what happens to a science which eschews the judgmental, the particular, the personal, and becomes entirely abstract and computational (Sacks, 1987, p. 20).

I hope that cognitive psychologists will take heed of Dr. Sacks' warning and see the opportunity that survey research offers to offset the present trend toward abstract computationalism.

References

1. Suchman, L. and Jordan, B., "Interactional Troubles in Face-to-Face Survey Interviews," JASA, Vol. 85, No. 409, pp. 232-253, 1990.
2. Jabine, T., Straf, M., Tanur, J., and Tourangeau, R. (eds.), Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines, Washington, D.C.: National Academy Press, 1984.
3. Barr, A. and Feigenbaum, E.A., (eds.) The Handbook of Artificial Intelligence. Stanford, CA: Heuristech Press, 2:184-192, 1982.
4. Simon, H., "The Steam Engine and The Computer: What Makes Technology Revolutionary," EDUCOM Bulletin, 22(1):2-5, 1987.
5. Sacks, O., The Man Who Mistook His Wife For A Hat, New York: Harper and Row, p. 20, 1987.

**Reports Available in the
Statistical Policy
Working Paper Series**

1. Report on Statistics for Allocation of Funds (Available through NTIS Document Sales, PB86-211521/AS)
2. Report on Statistical Disclosure and Disclosure-Avoidance Techniques (NTIS Document Sales, PB86-211539/AS)
3. An Error Profile: Employment as Measured by the Current Population Survey (NTIS Document Sales PB86-214269/AS)
4. Glossary of Nonsampling Error Terms: An Illustration of a Semantic Problem in Statistics (NTIS Document Sales, PB86-211547/AS)
5. Report on Exact and Statistical Matching Techniques (NTIS Document Sales, PB86-215829/AS)
6. Report on Statistical Uses of Administrative Records (NTIS Document Sales, PB86-214285/AS)
7. An Interagency Review of time-Series Revision Policies (NTIS Document Sales, PB86-232451/AS)
8. Statistical Interagency Agreements (NTIS Documents Sales, PB86-230570/AS)
9. Contracting for Surveys (NTIS Documents Sales, PB83-233148)
10. Approaches to Developing Questionnaires (NTIS Document Sales, PB84-105055/AS)
11. A Review of Industry Coding Systems (NTIS Document Sales, PB84-135276)
12. The Role of Telephone Data Collection in Federal Statistics (NTIS Document Sales, PB85-105971)
13. Federal Longitudinal Surveys (NTIS Documents Sales, PB86-139730)
14. Workshop on Statistical Uses of Microcomputers in Federal Agencies (NTIS Document Sales, PB87-166393)
15. Quality in Establishment Surveys (NTIS Document Sales, PB88-232921)
16. A Comparative Study of Reporting Units in Selected Employer Data Systems (NTIS Document Sales, PB-90-205238)
17. Survey Coverage (NTIS Document Sales, PB90-205246)
18. Data Editing in Federal Statistical Agencies (NTIS Document Sales, PB90-205253)
19. Computer Assisted Survey Information Collection (NTIS Document Sales, PB90-205261)
20. Seminar on the Quality of Federal Data (NTIS Document Sales, PB91-142414)

Copies of these working papers may be ordered from NTIS Document Sales, 5285 Port Royal Road, Springfield, VA 22161 (703) 487-4650